

**Improving ethnic data collection for statistics of cancer incidence, management, mortality and survival in the UK**

IQBAL, Gulnaz, GUMBER, Anil <<http://orcid.org/0000-0002-8621-6966>>, SZCZEPURA, Ala, JOHNSON, Mark, WILSON, Sue and DUNN, Janet

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/14818/>

---

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

**Published version**

IQBAL, Gulnaz, GUMBER, Anil, SZCZEPURA, Ala, JOHNSON, Mark, WILSON, Sue and DUNN, Janet (2007). Improving ethnic data collection for statistics of cancer incidence, management, mortality and survival in the UK. Project Report. University of Warwick for Cancer Research UK. (Unpublished)

---

**Copyright and re-use policy**

See <http://shura.shu.ac.uk/information.html>

# Improving ethnic data collection for statistics of cancer incidence, management, mortality and survival in the UK

**Gulnaz Iqbal<sup>1</sup>, Anil Gumber<sup>2</sup>, Ala Szczepura<sup>2</sup>,  
Mark RD Johnson<sup>2,3</sup>, Sue Wilson<sup>4</sup>, Janet A Dunn<sup>1</sup>**

<sup>1</sup> Warwick Medical School Clinical Trials Unit, University of Warwick

<sup>2</sup> Centre for Evidence in Ethnicity Health and Diversity, University of Warwick

<sup>3</sup> Mary Seacole Research Centre, De Montfort University, Leicester

<sup>4</sup> Department of Primary Care, University of Birmingham

Project awarded South Birmingham Research Ethics Committee approval  
(ref: 07/Q2707/33) on 26<sup>th</sup> March 2007

## Acknowledgements

Gulnaz Iqbal (née Begum) was part-funded by a Population and Behavioural Science grant provided by Cancer Research UK and the University Hospitals Coventry and Warwickshire Trust.

We would like to thank the Advisory Board (appendix 1) for their helpful comments on the progress and provisional study results, Diane Clay, the information officer for Evidence in Ethnicity Health and Diversity (CEEHD) who expertly carried out the electronic searches, Roger Gadsby for the use of his general practice database, the focus group facilitators as well as the focus group volunteers, Anurita Mulchand at 'Race for Health', the National Cancer Research Network (NCRN) for sending out the questionnaire and all responders for openly sharing their opinions, and finally Professor Bob Grieve and his team at the Arden Cancer Network for sharing their expertise and experience in collecting ethnicity data in a busy NHS setting.

This work was commissioned by Cancer Research UK and we would like to thank members of their policy department Paul Haezewindt, Ruth Yates, Charlotte Moore and Vanessa Gordon-Dseagu for their support.

## Executive summary

### Background

- Ethnic record keeping and monitoring has been identified as being important for over the past 20 years and has often been associated with negative rather than positive reasons
- There is an identified need for the collection of ethnicity in healthcare settings
- There is an “Evidence desert” with few “best practice reports”
- There have been recent initiatives to collect ethnicity through Hospital Admissions (e.g. Hospital Episode Statistics) and Primary Care

### Findings of this feasibility project:

#### Systematic review

- There is a paucity of published evidence regarding methodology of ethnicity data collection or ethnicity profiling
- Many clinical articles use ethnicity data in their analyses, health surveys or risk assessment of particular diseases, but often the data are of variable quality
- Self assessed ethnicity is better than observer assessment
- Official ethnicity categories need to be re-examined and modified to represent new migration patterns

#### Professional perception:

- Lack of consistency at different levels of organisations
- No clear rationale for collection/use of data
- Data collected without training or explanation of its use
- Not enough resources for ‘research data’ in general, without adding ethnicity

#### Focus groups perception:

- No objections to providing data in a healthcare setting
- Willingness to engage in research
- Lack of information as to the use of the data

#### Groups actively looking at how to collect and use ethnicity data

- South Thames registry looking at cancer rates for South Asians
- Healthy Londoners’ project and London Public Health Observatory initiatives
- Census categories extended to capture ‘British Born’ groups
- Specialist libraries and JISC-mail electronic discussion groups already set up

#### Recommendations – need for better training and information

- Need training materials/workshops to raise awareness
  - Patients- why your doctor should know your ethnic group?
  - Professionals- how to ask and explain the importance of ethnicity data collection?
- Work towards a culture of routine data collection of ethnicity at GP level
- Need working groups to assess collection, completeness and validation
- Need more reporting of ethnicity data in the healthcare setting, in order to improve planning and delivery of services for ethnic minority groups

## **Glossary**

CEEHD	Centre for Evidence in Ethnicity Health and Diversity
SLEH	Specialist Library for Ethnicity and Health
DoH	Department of Health
CR UK	Cancer Research UK
NHS	National Health Service
R&D	Research and Development
PCT	Primary Care Trust

## CONTENTS

1	Purpose .....	1
2	Background.....	1
3	Project objectives.....	7
4	Methods.....	8
4.1	Systematic literature review.....	8
4.1.1	Search criteria for published literature.....	8
4.1.2	Method of selection .....	8
4.1.3	Grey literature review .....	9
4.2	Ethnicity data collection questionnaire.....	10
4.3	Community participant focus groups .....	11
4.4	GP validation exercise .....	12
5	Results.....	13
5.1	Literature search.....	13
5.1.1	Published Literature.....	13
5.1.2	Grey literature.....	20
5.1.3	Summary of literature review.....	24
5.2	Health Professionals questionnaire .....	25
5.2.1	Do you attempt to collect any ethnicity data on patients?.....	25
5.2.2	Reasons given why ethnicity data are not collected .....	26
5.2.3	Disease areas for which ethnicity data are routinely collected ..	27
5.2.4	Method of collection.....	27
5.2.5	Indicators routinely collected .....	27
5.2.6	Problems encountered during ethnicity data collection.....	28
5.2.7	Perceived importance of ethnicity data collection.....	29
5.2.8	Summary .....	29
5.3	Patients' experiences and perceptions .....	30
5.3.1	General opinions on the collection of ethnicity .....	30
5.3.2	Experiences of providing ethnicity information .....	34
5.3.3	Categories used in practice .....	37
5.3.4	Language, religion and culture .....	38
5.3.5	How information should be collected? .....	40
5.3.6	Overall summary .....	42
5.4	Validation of General Practice data .....	43
6	Synthesis of the results of the project.....	44
6.1	Ethnicity data collection and monitoring .....	45
6.2	Categories for defining ethnic group.....	47
6.3	Other Indicators of ethnicity.....	48
6.4	Methods of collection .....	49
6.5	Barriers to collection .....	51
6.6	Interventions.....	52
6.7	Data quality and completeness.....	54
7	Conclusions .....	55
8	References .....	56
	Appendix 1: CanEth and Advisory board members.....	59
	Appendix 2: Search terms for literature search .....	60
	Appendix 3: Ethnicity, Health and Diversity questionnaire .....	61
	Appendix 4: Topic guide.....	63
	Appendix 5: NCRI poster 2007.....	64

## 1 Purpose

- To gather robust evidence and identify clear solutions to improve the collection of ethnicity data for cancer. Evidence and recommendations will be used to inform CR UK policy makers regarding ethnic cancer statistics and be used to lobby the Government to improve ethnic cancer data collection.
- Recommendations from this project could also feed into an updated Cancer Plan, where health inequalities and the needs of minority ethnic groups are likely to feature prominently.

## 2 Background

Cancer Research UK (CR UK) identified the need to collect accurate data on 'ethnicity' and recognised this as being essential to inform policy makers, funders and public health experts on the incidence, prevalence and outcomes of specific conditions in sub-groups of the population. There is evidence that certain ethnic groups have an increased incidence of some conditions (e.g. diabetes, hypertension, stroke and certain cancers), and that there are disparities in access to services [1]. The accurate collection of 'ethnicity' data can help to 1) quantify the size of the problem, 2) assist with the design of appropriate interventions, 3) allow for education, management and resources to be appropriately targeted to high risk populations, and 4) assess the benefit of such interventions with regards to reducing the disease burden and improving mortality.

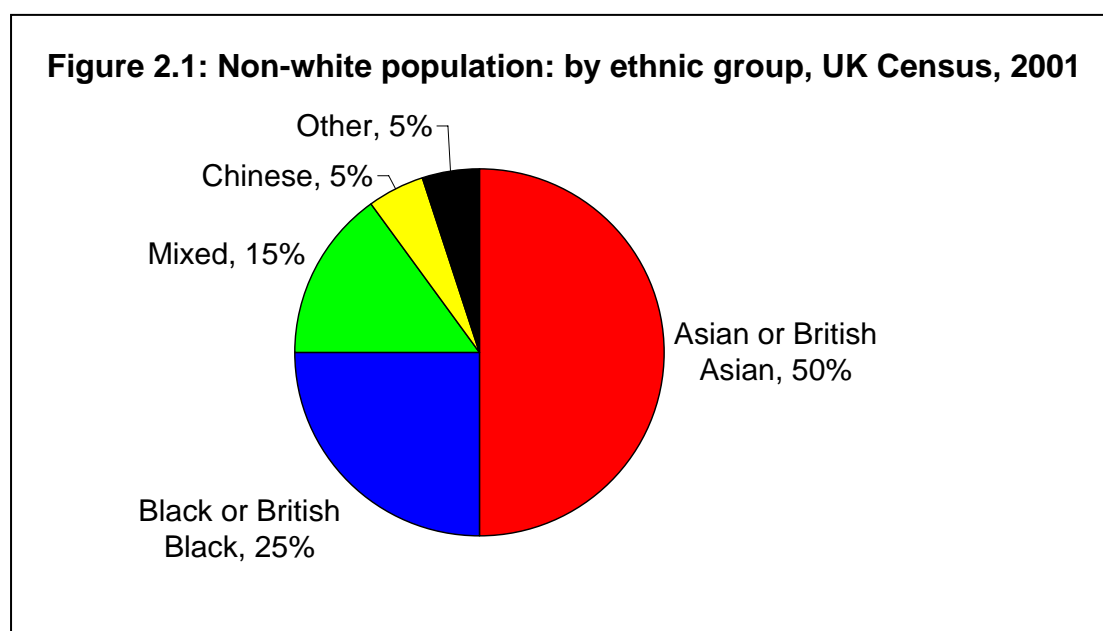
In cancer, ethnicity data collection and monitoring is important because ethnic minorities are associated with later presentation leading to poor survival [2-4]. Also some ethnic minority groups are associated with more risky behaviour; e.g. smoking rates were reported to be the highest in Bangladeshi males (44%) followed by Irish males (39%), compared to 27% in the general population, whereas Bangladeshi women are more likely to chew tobacco (26%) than smoke cigarettes [2]. Reports have suggested that both breast and colorectal cancer incidence were lower in the South Asian population, however incidence rates are increasing over time [5, 6].

In other disease areas, South Asians in the UK are 50% more likely to die prematurely from coronary heart disease than the general UK population; Males and females of Pakistani and Bangladeshi origin are 6 times more likely than the general population to have diabetes; Women of Caribbean origin are most likely to be obese, with 50% of those aged 55+ falling into this category [7, 8]. In an attempt to gather evidence on inequalities in health, two major reports published in the early 1980's were updated and presented in one volume in 1988 [8]; the joint publication being the key resource on health inequalities for the past 20 years. These reports present some inequalities by ethnic group, as well as income and social class. The 1976 UK Race

Relations Act was amended in 2000, adding obligations on NHS bodies to promote equality of access and quality of healthcare services in a multicultural society. In the USA, the 1993 Revitalization Act set a target that all randomised controlled trials must include at least 30% of minorities to be recruited as previously there was an under representation of minorities being offered the benefit of being treated within a trial.

The definition of ethnicity is complex. It has been defined by Bhopal (2004) as being “the social group a person belongs to, and either identifies with or is identified with by others, as a result of a mix of cultural and other factors including language, diet, religion, ancestry and physical features” [9]. Ethnicity is often associated with race, which is more dependent on geographical origins and physical features. In the USA, ethnic/racial group are collected as two separate fields (NIH policy, 2001) [10]. In the UK, ethnic group is collected as one field with elements of ethnicity and race combined.

The 2001 census classified 4.6 million people (7.9%) in the UK as belonging to a ‘non-white’ ethnic group (UK census, 2001) [11] with over half of these classified as ‘Asian’ or ‘British Asian’, figure 2.1. This is an increase compared to the 5.5% of the population not defined as ‘white’ from the 1991 census.

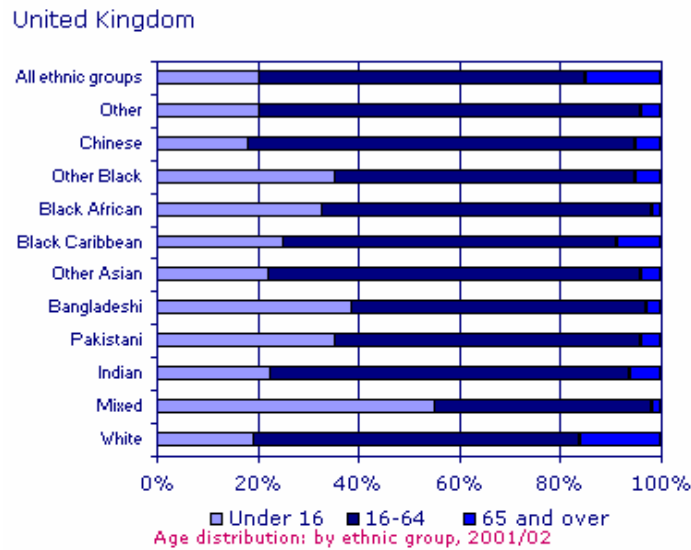


There are differing views on the reasons for variability in incidence rates between ethnic groups, with some suggesting genetic or environmental (e.g. diet) differences may be responsible, and others that it could be due to immigrant populations having lower incidences because of the younger age structure than the white majority, Figure 2.2 [11]. There is a significantly lower percentage of Black African, Mixed and South Asians who are over 65 years old when compared to White group. Since cancer is predominantly a disease of the elderly, the age structure differences could explain lower cancer incidence, in certain minority groups.



**Figure 2.2: Age distribution by ethnic group, 2001/2002**

([http://www.statistics.gov.uk/downloads/theme\\_social/social\\_focus\\_in\\_brief/ethnicity/ethnicity.pdf](http://www.statistics.gov.uk/downloads/theme_social/social_focus_in_brief/ethnicity/ethnicity.pdf))



The 2001 census identified 55% of the 'mixed race' category as being 16 year olds or younger, (figure 2.2). For epigenetic modelling, a more detailed definition of 'mixed' is required such as mothers, fathers and grandparents ethnicity and geographical origins/ancestry. For removal of barriers in health care systems with respect to ethnicity, there is a need to break down 'ethnicity' further to identify language, religion and culture, thus allowing more accurate information to be collected and resources to be optimally targeted.

In the UK the ethnicity debate has often focused on the utility and classification of ethnicity data [2, 12-16]. The quality of ethnicity data recording has been highly variable and attempts to improve the completeness and quality frequently resisted. Reports focussing on ethnicity tend to use the standard Census categories but frequently show significant numbers of cases reported as 'not known' or 'did not answer question' and consequently the impact and value of such work is limited [2, 13]. Recording of additional dimensions of diversity, such as religion or preferred language, is infrequent and often poorly conducted.

In general, collection of ethnicity data has long been recognised as poor in the UK, especially in primary care, with respect to completeness and accuracy [17-19]. There are many reasons for the lack of routinely collected ethnicity data. These include the difficulty of an accurate classification, awareness of sensitivities when asking for these data, lack of motivation to collect or provide data, unwillingness (or inability due to language barriers) for individuals to provide information, and lack of understanding as to how such data can or will be used. Reports on health inequalities and outcomes across ethnic groups emphasise the need to overcome these barriers and accurately record ethnicity; the danger is that current policies are based on inaccurate data and, as such, may lead to inadequate resources and services [2, 12, 13].

In 1995 it became UK Government policy to collect ethnicity in a secondary care setting through HES. HES data collection has improved in London since the first collections in 1996/1997 from 52% of records with incomplete data to 35% in 2001/2002 [12].

In 2002, an attempt was made to increase ethnicity profiling in primary care; however, at this time the work involved and associated costs were significant deterring factors [18]. Recently some PCTs have invested in the collection of ethnic group and these initiatives are supported by the incorporation of ethnicity into the Quality and Outcomes Framework for GPs (although restricted to new patients and only awarded 1 point). Monitoring goals set for London 2003-2006 by the Department of Health expected all GP practices and other primary care providers to record valid ethnicity codes for 75% of patients by 2005 and reaching 95% by March 2006 [12]. The "Professionals Responding to Cancer in Ethnic Diversity" (PROCEED) project team provided training in competence and cultural awareness for health professionals involved in cancer care at the primary care level. The issues explored include 'Cancer and Ethnic Diversity', 'Language and Communication' and 'Culture and Cancer' [20].

In the absence of robust self-assigned ethnicity data, name analysis can offer a useful alternative for the identification of South Asian (the largest minority group) and other populations with distinct names [21-23]. CEEHD (The UK Centre for Evidence in Ethnicity, Health and Diversity based at Warwick Medical School) has undertaken comparisons of the two main name recognition software packages currently available (Nam Pehchan and SANGRA) using local datasets containing self-assigned ethnicity. Nam Pehchan provided higher sensitivity and positive predictive values for the West Midlands population when compared to SANGRA [24, 25]. The use of the Nam Pehchan software has previously been shown to be reliable (91% sensitivity, 99% specificity) [21]. The technique has been used successfully to estimate ethnic uptake of bowel cancer screening in the Coventry & Warwickshire pilot site [26, 27]. A survey of all PCTs in England, carried out by CEEHD in 2004/2005 indicated that only 1 in 10 PCTs have access to this software. A literature search identified 54 papers reporting use of name recognition software [25]. Only one study reported use in general practice, solely for determining indicative prevalence of diabetes by practice [28].

In 2005 the NHS produced a 'practical guide to ethnic monitoring in the NHS and social care' with several examples of good practice [29]. There is limited information on the uptake of these guidelines and their practical applicability. Within the cancer setting, family history, social class, material deprivation, lack of access to services and subsequent delay times have all been adversely linked to outcome (i.e. survival) [2, 4, 6, 8]. There is an urgent need for evidence on how ethnic data collection might be improved for cancer statistics, what mechanisms might be implemented for data quality validation checks, and a strategy for optimal use of this data in order to encourage improved collection.

## Background References

1. Atkinson M, et al., *Systematic review of ethnicity and health service access for London. Report to London Regional Office NHS Executive. ISBN0-9535-4300-6*. 2001, Coventry: Centre for Health Services Studies, with MSRC and CRER, for London Regional Office, NHS Executive.
2. White A, *Social Focus in brief: Ethnicity*. 2002, Office for National Statistics. London.
3. Smith LK, et al. *Incidence of Cancer in Leicestershire: The Effects of Ethnicity and Deprivation*. in *International Association for Critical Realism*. 1999. Lisbon, Portugal.
4. Woods, L.M., B. Rachtel, and M.P. Coleman, *Origins of socio-economic inequalities in cancer survival: a review*. *Annals of Oncology*, 2006. **17**(1): p. 5-19.
5. Smith LK, et al., *Latest trends in cancer incidence among UK south Asians in Leicester*. *British Journal of Cancer*, 2003. **89**(1): p. 70-73.
6. Farooq, S. and M.P. Coleman, *Breast cancer survival in South Asian women in England and Wales*. *Journal of Epidemiology and Community Health*, 2005. **59**(5): p. 402-406.
7. *Statistics: Health care services*. [cited; Available from: [www.cre.gov.uk/research/statistics\\_health.html](http://www.cre.gov.uk/research/statistics_health.html)].
8. Townsend P, Davidson N, and Whitehead M, *Inequalities in health: the Black report and the Health Divide*. 1988: Penguin Books.
9. Bhopal, R., *Glossary of terms relating to ethnicity and race: for reflection and debate*. *Journal of Epidemiology and Community Health*, 2004. **58**(6): p. 441-445.
10. *NIH policy on reporting race and ethnicity data: subjects in clinical research 2001* [cited; Available from: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-01-053.html>].
11. *UK Census 2001*. [cited; Available from: <http://www.statistics.gov.uk/census2001/census2001.asp>].
12. *Diversity counts. Ethnic health intelligence in London: the story so far*. 2003, London Health Observatory.
13. Mackintosh M, *London - the world in a city. An analysis of 2001 Census results*. 2005, Data Management and Analysis Group.
14. Johnson MRD, *Ethnic monitoring: bureaucratic construction of a minority entity or identity*, in *Social Work and Minorities*, Williams C, Soydan H, and Johnson M, Editors. 1998, Routledge. p. 79-90.
15. Johnson MRD, *Ethnicity*, in *Public health evidence: Changing the health of the public Oxford*, Killoran A, Swann C, and Kelly M, Editors. 2006, University Press.

16. Johnson MRD, *Ethnic Monitoring and Nursing in Ethnicity and Nursing Practice*, Culley L and Dyson S, Editors. 2001, Macmillan/Palgrave. p. 91-106.
17. Pringle, M. and I. Rothera, *Practicality of recording patient ethnicity in general practice: Descriptive intervention study and attitude survey*. British Medical Journal, 1996. **312**(7038): p. 1080-1082.
18. Jones M and Kai J, *Ethnicity profiling in primary care: A pilot study in Midlands Research Practices Consortium*. 2003, University of Birmingham.
19. Kumarapeli, P., et al., *Ethnicity recording in general practice computer systems*. Journal of Public Health, 2006. **28**(3): p. 283-287.
20. *The PROCEED project*. [cited; Available from: <http://info.cancerresearchuk.org/proceed/>].
21. Cummins, C., et al., *An assessment of the Nam Pehchan computer program for the identification of names of south Asian ethnic origin*. Journal of Public Health Medicine, 1999. **21**(4): p. 401-406.
22. Nanchahal, K., et al., *Development and validation of a computerized South Asian Names and Group Recognition Algorithm (SANGRA) for use in British health-related studies*. Journal of Public Health Medicine, 2001. **23**(4): p. 278-285.
23. Macfarlane, G.J., et al., *Determining aspects of ethnicity amongst persons of South Asian origin: The use of a surname-classification programme (Nam Pehchan)*. Public Health, 2007. **121**(3): p. 231-236.
24. Honer D, *Identifying Ethnicity: A comparison of two software programmes designed to identify names of South Asian origin*. 2003, MPH Dissertation, University of Birmingham.
25. Gumber A. *Ethnicity and Name Analysis: Potential for Improved Ethnic Monitoring in the NHS Using Nam Pehchan*. in *Royal Geographical Society Annual International Conference on Global Social Justice and Environment Sustainability*. 2006. London.
26. Szczepura A, et al., *Ethnicity: UK Colorectal Cancer Screening Pilot*. 2003, Department of Health.
27. Gumber A, Szczepura A, and Charlotte B. *Improving Equity and Efficiency in Cancer Screening Programmes*. . in *Fourth Annual Meeting of Health Technology Assessment International (HTAi) on HTA for Evidence-Based Public Health*. 2007. Spain.
28. Ainley C and Corcoran M. *Baseline survey of primary care services for people with diabetes in Nottingham health district*. in *Trent Public Health Conference*. 2002. Nottingham.
29. *A Practical Guide to Ethnic Monitoring in the NHS and Social Care*. 2005, Department of Health and Social Care Information Centre/NHS Employers.

### **3 Project objectives**

1. Literature review of methods, interventions and barriers addressing the collection of ethnicity data (or ethnicity profiling) in primary and secondary care; including examples of good practice
2. Evaluation of health care professionals' perceptions and experiences of collecting ethnicity data in primary and secondary care
3. Evaluation of consumers' perceptions, experiences and willingness to provide ethnicity data in primary and secondary care
4. Validation (completeness and accuracy) of ethnicity data collected in a feasibility study of selected PCTs
5. Provision of a written report and presentation pack summarising the literature review, surveys and interviews, together with recommendations and examples of good practice

## 4 Methods

### 4.1 Systematic literature review

The literature review encompassed searches of published literature on bibliographic databases supplemented with World Wide Web Google searches and searches of specific websites to identify “grey” literature.

#### 4.1.1 Search criteria for published literature

Literature searches were carried out using Embase, Psychlit, MEDLINE, PsycINFO and CINAHL by a specialist ‘ethnicity’ information scientist.

The three key search areas were: “ethnicity”, “data collection” or “data monitoring” and “cancer” or other chronic or long term diseases such as stroke, diabetes, coronary heart disease (see appendix 2 for a full listing of search terms). The first literature search was limited to 2000-2007 with the aim of identifying recent literature and up to date good practice guidelines. The second literature search used the same search terms as the first but was limited to literature published 1990-1999; the National Institute of Health Revitalisation Act was passed in the USA in 1993 prompting interest in ethnic groups. This search was done to ensure all relevant published literature was included in the review.

The third literature search was cancer site specific with no set time limitation. The five main sites were Breast, Colorectal, Prostate, Head and Neck and Cervical. Search terms were used as described above and also included cancer sites. Site specific searches were conducted in each database uniquely per disease site. These were then concatenated. Duplicates were removed for each disease site using the Endnote software and then further by hand.

#### 4.1.2 Method of selection

The review process was broken down into 3 stages:

Stage 1: Titles were reviewed and coded as “not relevant”, “possibly relevant” or “definitely relevant”. Abstracts were printed for the titles coded as “possibly relevant” or “definitely relevant” for the second stage.

Stage 2: Abstracts were reviewed and coded as:

**M**=Definitely/possibly contains some methodology of ethnicity data collection/monitoring

**U**=Use of ethnicity data but no explanation of how the data was collected

**I**=Insubstantial

**D**=Duplicate

**N**=Not relevant

Abstracts coded as “Definitely/possibly contains some methodology of ethnicity data collection/monitoring” proceeded to the final stage.

Stage 3: These articles were read, coded and summarised in the minimum data extraction table.

**M**=Definitely/possibly contains some methodology of ethnicity data collection/monitoring

**U**=Use of ethnicity data but no explanation of how the data was collected

**I**=Insubstantial

**N**=Not relevant

Only the relevant methodology articles are summarised in the results table.

### **4.1.3 Grey literature review**

Grey literature searches were conducted using the keywords in the search as follows: (“data collection” OR “data monitoring”) AND (“ethnic” OR “ethnicity”).

The searches were performed in Google and Google Scholar. Due to the huge volume of results using this method, the reviewer scanned the results of the first 50 pages only for each search. The majority of articles were rejected on the grounds that they discussed the need for the collection of ethnicity or used ethnicity data for reporting outcomes. Only articles which described their own ethnicity data collection policies or procedures, and those giving guidance were considered.

In addition, extensive searches were carried out on key websites and links from these websites:

- Specialist Library for Ethnicity and Health (SLEH)  
(<http://www.library.nhs.uk/ethnicity/>)
- Centre for Evidence in Ethnicity Health and Diversity (CEEHD)  
(<http://www2.warwick.ac.uk/fac/med/research/csri/ethnicityhealth/>)
- London Health Observatory (LHO)  
(<http://www.lho.org.uk/>)
- National Cancer Library  
(<http://www.library.nhs.uk/cancer/>)
- Office for National Statistics (ONS)  
(<http://www.statistics.gov.uk/>)
- Department of Health  
(<http://www.dh.gov.uk/en/index.htm>)

Findings from relevant grey literature were summarised in the results table.

## 4.2 Ethnicity data collection questionnaire

The questionnaire was based on one previously developed by the Centre for Evidence in Ethnicity Health and Diversity (CEEHD) and modified by the CanEth working group for the project. The modified two page questionnaire comprised of nine items including perceived importance, current practice, reasons for not collecting ethnicity data, problems encountered when data are collected, disease areas, method of collection, items collected, use of name recognition software and ethnicity data collection training (appendix 3). The questionnaire was designed to be quick and easy to complete, ensuring the majority of questions could be answered using tick box responses with additional space for comments. The questionnaire could be printed, completed and returned by post or sent by email.

The questionnaire was aimed at clinicians, managers and nurses and anyone else who may be involved in collecting or using ethnicity data in a healthcare setting (e.g. statisticians, information scientists, data managers). Questionnaires or a link to the questionnaire were distributed to:

- Minority-Ethnic-Health jiscmail list  
(<http://www.jiscmail.ac.uk/lists/MINORITY-ETHNIC-HEALTH.html>)
- ALLSTAT jiscmail list, emailed  
(<http://www.jiscmail.ac.uk/lists/allstat.html>)
- Emailed to National Cancer Research Network (NCRN) head office for circulation
- Questionnaire posted on CEEHD website  
(<http://www2.warwick.ac.uk/fac/med/research/csri/ethnicityhealth/>)
- Link to questionnaire on CEEHD posted on SLEH website  
(<http://www.library.nhs.uk/ethnicity/>)
- Thread created on Academic Clinical Oncology and Radiobiology Research Network (ACCORN) thread  
(<http://www.acorn.org/ResearchDB/>)
- Thread created on NHS discussion forum and news item on `new@networks` electronic bulletin in June  
(<http://www.networks.nhs.uk/forums/>)
- News item in the Wales Cancer newsletter
- Emailed to all Cancer Network managers in England and Wales
- Circulated to all Race for Health PCTs, also posted on `Race for Health` website <http://raceforhealth.org/>

A 4 week deadline to return the questionnaire was set and extended for a further 4 weeks (on the web-site links) to increase response. The NCRN was the only mailshot to be repeated after an initial poor response, when sent through the NCRN head office. The repeated mailshot was sent individually to each network manager which improved the response rate.



### 4.3 Community participant focus groups

Focus groups were conducted in collaboration with the Mary Seacole research centre at De Montfort University and the Ethnic Health Forum in Manchester. A topic guide (appendix 4) was developed by the CanEth group, together with information sheets and consent forms. Ethical approval was required and obtained through the South Birmingham LREC.

Focus groups were carried out by trained facilitators who recruited the volunteers from local community centres and places of worship. The focus groups undertaken were dependant upon the availability of facilitators who were required to speak in the native tongue of the focus group, in this case:

- Mirpuri speaking Muslim females (Azad Kashmir) – facilitator 1
- Bengali speaking Muslim males (Bangladeshi) - facilitator 2
- Urdu speaking Muslim males (Pakistani) - facilitator 3
- Urdu speaking Muslim females (Pakistani) - facilitator 4
- Punjabi speaking Hindu males and females (Indian) - facilitator 5

Informed consent was taken by the facilitator where English was not the volunteers' preferred language. A brief introduction was given by the facilitator in the language appropriate for the group. Groups ranged from 5 to 10 people, seating was arranged in a circular formation to encourage discussion. Often the discussion took place in both the native language and English; this was particularly the case when including younger volunteers.

The facilitators used the topic guide which was specifically developed to focus on the five areas of interest:

- General opinions on the collection of ethnicity information
- Experiences of providing ethnicity information
- Categories used in practice
- Language, religion and culture
- How information should be collected

This was followed by general comments and closing remarks.

All sessions were recorded and transcribed by the facilitators who summarised the findings in the form of a bullet point report and were analysed by researchers at the University of Warwick.

## 4.4 GP validation exercise

Applying the 'Nam Pehchan' name recognition software to local general practice databases will allow identification and validation of South Asian patients.

### Validation Standard Operating Procedure given to GPs:

**Key aim:** Validation (completeness and accuracy) of ethnicity data collected in a feasibility study of selected PCTs. The name recognition software 'Nam Pehchan' will identify names of South Asian origin and their corresponding ethnic group. Accuracy of existing PCT ethnicity data will be assessed by comparing South Asian ethnicity in the Practice dataset to the results from the 'Nam Pehchan' analyses.

**Time frame:** The exercise should not take more than a few hours

### Procedure as follows:

1. Practice staff will create a dataset containing: **patient forename, middle name, surname, date of birth, place of birth, date of registration and ethnicity information fields** for all patients and save temporarily onto a pen drive
2. The pen drive will be opened on the University of Warwick laptop which will have the name recognition software 'Nam Pehchan' pre-installed
3. The dataset will be analysed using the Nam Pehchan' program
4. Once the 'Nam Pehchan ethnicity' and the 'practice ethnicity' have been compared, summary data will be tabulated and saved onto the laptop
5. The pen drive containing practice data will be wiped clean and removed, unless the practice would like to keep a copy of the 'Nam Pehchan ethnicity field'
6. The summary results will be analysed at Warwick and fed back to the practice

R&D approval was obtained for South Birmingham, Coventry and Warwickshire. Four research active practices were approached for permission to apply the name recognition software to their databases and to compare these results with their existing ethnicity field. One of these practices allowed access to their database within the timescale of this project.

## 5 Results

### 5.1 Literature search

The systematic review was carried out in two parts. The first part used the search criteria in appendix 2 to look at published literature between 1990-1999, 2000-2007 and cancer specific sites. The second part looked at grey literature pulled from key websites as well as Google searches. Both parts are assessed using different criteria; the published papers assessed in three stages – 1) title review, 2) abstract review and 3) article review, and the grey literature reviewed by full article. The number of results from the grey literature searches are too many and too duplicated to be able to present rejected articles in a meaningful way, hence only the relevant articles are presented.

#### 5.1.1 Published Literature

The systematic review of the published literature provided a total of 2404 'hits'; 720 for the period 1990-1999 and 1684 for 2000-2007 (table 5.1).

**Table 5.1: Search 'hits' by database**

Literature search	Database	No of 'Hits'
1990-1999	Medline	492
	PsychInfo	147
	CINAHL	64
	Embase	380
	Duplicates	362
	<b>Total remaining</b>	<b>720</b>
2000-2007*	Medline	1059
	PsychInfo	356
	CINAHL	96
	Embase	173
	<b>Total remaining</b>	<b>1684</b>
	<b>Grand Total</b>	<b>2404</b>

\* Numbers excluding duplicates are shown (total considered 2658)

Upon review of the 2404 titles, only 322 seemed to suggest that they involved the methodology of either collecting or monitoring ethnicity data. A full review of these 322 abstracts revealed only 26 which potentially fulfilled our criteria, table 5.2. The main reason for rejection (57%) was that the paper was concerned with 'use' of ethnicity rather than the methods for collection.

**Table 5.2: Stages 1 and 2: Title and abstract review**

Period	No of titles	No of abstracts	Abstract category*					No of articles
			M	U	I	D	N	
1990-1999	720	218	8	127	46	7	30	8
2000-2007	1684	104	18	57	29	0	0	18
Total	2404	322	26	184	75	7	30	26

\*M=Methodology, U=Use, I=Insubstantial, D=Duplicate, N=Not relevant

The full text of the 26 potential articles which assessed ethnicity data collection or ethnicity monitoring, where reviewed; only 19 included information about data collection or monitoring. Again the main reason for rejection was a discussion on 'use' of ethnicity and not data collection or monitoring (table 5.3). One of the potentially relevant papers was included based on the abstract only as the full paper was unavailable at the time of review.

**Table 5.3: Stage 3 article review**

Period	No of articles reviewed	Article category*				No of articles remaining
		M	U	I	A	
1990-1999	8	7	1	0	0	7
2000-2007	18	11	4	2	1	12
Total	26	18	4	2	1	19

\*M=Methodology, U=Use, I=Insubstantial, A=Abstract only

Due to the low numbers of papers fulfilling the criteria for acceptance and the interest in cancer sites, the search was repeated for specific cancer sites as in table 5.4.

**Table 5.4: 'Hits' by database for cancer site specific searches**

Literature search	Database	No of 'Hits'
Breast	Medline	151
	PsychInfo	60
	CINAHL	24
	Embase	119
	Duplicates	102
	Total remaining	252
Colorectal	Medline	71
	PsychInfo	4
	CINAHL	5
	Embase	63
	Duplicates	38
	Total remaining	105
Cervical	Medline	50
	PsychInfo	26
	CINAHL	20
	Embase	39
	Duplicates	41
	Total remaining	94
Prostate	Medline	42
	PsychInfo	4
	CINAHL	2
	Embase	43
	Duplicates	24
	Total remaining	67
Head and neck	Medline	15
	PsychInfo	1
	CINAHL	2
	Embase	12
	Duplicates	9
	Total remaining	21

This resulted in 539 potential articles of which 469 were deemed to possibly fulfil the criteria for acceptance after review of the title as shown in table 5.5.

**Table 5.5: Cancer site specific Stages 1 and 2 review**

Period	No of titles	No of abstracts	Abstract category*					No of articles
			M	U	I	D	N	
Breast	252	231	0	111	38	7	75	0
Colorectal	105	87	2	51	1	3	30	2
Cervical	94	81	1	52	15	0	13	1
Prostate	67	53	0	37	9	0	7	0
Head & Neck	21	17	1	4	0	0	12	1
Total	539	469	4	253	62	10	137	4

\*M=Methodology, U=Use, I=Insubstantial, D=Duplicate, N=Not relevant

Only 4 articles out of the 469 abstracts reviewed involved ethnicity data collection or monitoring, and 2 of these were duplicates from the previous non-disease site specific searches. The other 2 papers did not involve data collection or monitoring upon full review of the paper, table 5.6.

**Table 5.6: Cancer site specific Stage 3 article review**

Period	No of articles	Article category*				No of articles remaining
		M	U	I	A	
Breast						
Colorectal	2		1		1	1 duplicate
Cervical	1		1			
Prostate						
Head & Neck	1	1				1 duplicate
Total	4	1	2		1	2

\*M=Methodology, U=Use, I=Insubstantial, A=Abstract only

In summary for the literature review of the published literature, 19 papers were identified for acceptance (tables 5.7 and 5.8).

**Table 5.7: Extracted from all selected articles- 2000 to 2007**

ID	Author , year Of publication	Type of cancer	Country of study	Ethnic group	Type of study	Description of content	
						Focus(es) of study	Key findings
1	Baker, 2007	Non- cancer specific	USA	All	Cross sectional	Patients attitudes towards healthcare providers collecting their ethnicity, race and language data	88% of patients thought the data should be collected. 46% worried that the information would be used to discriminate against them. 17% were not comfortable reporting their own ethnicity.
2	Ma, 2007	Non- cancer specific	All	All	Systematic review	Methods of reporting race in medical journal articles	116 terms used to describe ethnic groups, only 13% reported data collection method (1152 articles)
3	Weinick, 2007	Non- cancer specific	USA	All	Review	New enactment of ethnicity data collection in acute care hospitals. Lessons learnt from implementing publicly mandated data collection	Implementation of a change of policy needs to map onto existing systems, be flexible and be standardised. Training the trainer central sessions proved successful. Patient engagement and emphasis on the importance of data collection for improvements of care.
4	Hasnain-Wynia, 2006	Non- cancer specific	USA	All	Overview	Ethnicity data collection in healthcare, current practice, barriers and solutions	Highlighted the need for self- reporting, why the data are needed and how professionals should ask for it
5	Jack, 2006	All	UK	All	Audit	To determine completeness of ethnicity data in Thames cancer registry and HES data held by London Health Observatory	81% of HES data had ethnicity recorded compared to 23% in the registry. Better collaboration needed between sources in order to improve registry ethnicity data

6	Baker, 2005	Non-cancer specific	USA	All	Cross sectional	Patients attitudes towards healthcare providers collecting ethnicity data	Patients more willing to provide ethnicity data when reasons for collection are explained by staff in an appropriate manner. Staff should be comfortable collecting this data
7	Buescher, 2005	Live birth records	USA	All	Audit	Discrepancies between published data on racial classification and self-reported race	Measures of racial disparity vary depending on whether self-reported or official coded race is used
8	Ford, 2005	Veteran Affairs	USA	All	Review	Importance of conceptualising and categorising ethnicity data	Better and more consistent methods of ethnicity data collection need to be developed
9	Gotay, 2004	All	Hawaii	Japanese Hawaiian Europeans Filipinos	Cross sectional	To assess ethnic self-identity in 367 recently diagnosed ethnic patients. Explores acculturation.	Findings show medical records well linked to individual self-reported ethnicity
10	Lin, 2001	All	USA	All	Audit	SEER initiative to assess the completeness of country of birth data	67% of patients on the register had birthplace recorded. Completeness varied between ethnic groups suggesting bias in collection of this item
11	Chattar-Cora, 2000*	Colorectal	USA	All	Audit	To determine the demographic and tumour characteristics of a multi-ethnic group	Patient notes were used to successfully identify 685 out of 688 patients. Ethnicity could not be identified for 3 patients
12	Olatokunbo, 2000	Non-cancer specific	UK	All	Feasibility study	Ethnic monitoring in primary care	Ethnic monitoring is feasible in primary care

**Table 5.8: Extracted from all selected articles- 1990 to 1999**

ID	Author, year of publication	Type of cancer	Country of study	Ethnic group	Type of study	Description of content	
						Focus(es) of study	Key findings
13	Centers for disease control, 1999	Non-cancer	USA	All	Report	To assess the collection of race data in health surveillance systems between 1994-1997	No improvement in race data collection was observed between 1994 and 1997
14	Warnakulasuriya, 1999	Mouth Pharynx Nasopharynx	UK	Asian Chinese	Audit	Incidence of head and neck cancers in Asian and Chinese groups, flagged by Thames cancer registry using name and place of birth	Ethnic groups can with certain precision be identified using names and place of birth data, as well as manual checking
15	Sheth, 1997	Non-cancer, Mortality database	Canada	South Asian Chinese	Audit	Novel method to identify ethnic origin using names and country of birth	Use of name and country of birth more accurate than using country of birth alone
16	Swallen, 1997	All cancer	USA	Hispanic	Audit	Misclassification of Spanish ethnic groups in cancer register using Census Spanish surname list, GUESS (name recognition software) and telephone interviews	This sample showed Hispanics over reported for 38% of cases. Recommends using both recorded ethnicity and name for increased accuracy
17	Kelly, 1996	Non-cancer, AIDS	USA	All	Audit	Validation of ethnicity classification for AIDS patients across 3 national data sources	Inconsistencies greatest for American Indians and Alaska Natives, up to 57% disagreement



18	Frost, 1994	Non-cancer	USA	American Indians Alaska Natives	Audit	To validate race on Washington death certificates with those in the Indian Health Service	Race was correct for 87% of death certificates. Deaths from cancer were more likely to be coded incorrectly. People who are born and died in Washington are more likely to be coded correctly
19	Sugarman, 1993	Non-cancer, End stage renal disease	USA	American Indians Alaska Natives	Audit	Misclassification of American Indians and Alaska natives in Renal disease stage register and impact upon disease statistics	Ethnicity validated against the Indian Health Service data using names, date of birth and social security numbers. Incidence of renal disease increased from 268 per million to 312 per million after corrections to ethnicity coding

### 5.1.2 Grey literature

Searches on key websites and Google identified a wealth of information linked to ethnicity. Upon review, 53 reports were identified as being possibly associated with ethnic data collection or monitoring. The main reason for rejection of possible reports was on the grounds that they discussed the need for the collection of ethnicity or used ethnicity data for reporting outcomes. Of the 53 reviewed, 16 reports were included in the review, table 5.9.

**Table 5.9: Grey literature summary table**

Id	Title, year of publication	Authors	Country of report	Ethnic groups	Type of research	Description of content	
						Focus(es) of report	Key findings
20	HRET Disparities Toolkit: A toolkit for collecting race, ethnicity and primary language information for patients (amended version), 2007	Health Research and Education Trust	USA	All	Online toolkit	Designed to help health care workers understand the importance of collecting good quality ethnicity, race and preferred language data	Toolkit includes topics: 1. Who should use the Toolkit 2. Why collect race, ethnicity, and primary language data 3. Why collect data using a uniform framework 4. The nuts and bolts of data collection 5. How to ask questions about race, ethnicity, and primary language 6. How to use race, ethnicity, and primary language data to improve quality of care 7. How to train staff to collect this information

21	Lambeth PCT review, 2006	Race for Health	UK	All	Paper	How successful is Lambeth Primary Care Trust at collecting, recording, analysing and using ethnicity monitoring information?	Good practice includes: 1. Individual Patient Registration Profile (IPRP) started in 2002 now over 30 practices are taking part. IPRP includes collection of religion, language, need for interpreter and ethnicity as well usual data. Existing patients contacted by means of postal questionnaire 2. Training for practice staff 3. Datatet system aids use of collected data
22	Race, ethnicity, and language of patients. Hospital practices regarding collection of information to address disparities in health care, 2006	Regenstein and Sickler, The Robert Wood Johnson Foundation	USA	All	Surveys	Current practices of US hospitals, completeness of data, methods of collection and barriers	Overall collection of data is good but not put to use. Some confusion between ethnicity and race. Single most important barrier to collection is staff not knowing why the data is important. Examples of good practice include: training given to new staff members as part of induction. Training for all staff collecting data on the importance of self-identification and uses of data, members of staff working in registration areas are subjected to quality review. Managers able to identify staff who record a large number of unknowns or blanks
23	Black and minority ethnic groups	Gill, Kai, Bhopal, Wild	UK	All	Needs assessment	A needs assessment overview for Black Minority Ethnic Groups (BMEGs) in the UK. Part of needs assessment series	No differences reported in the rate of minority groups consulting their GPs or been admitted to hospital. However, Afro-Caribbean males are less likely to have registered with a GP. Despite being mandatory there is still a lack of good ethnic data in secondary care services

24	A practical guide to ethnic monitoring in the NHS and social care, July 2005	Department of Health	UK	All	Guidelines	Practical guide to ethnic monitoring in the NHS	Examples of best practice in the UK, including self-reporting and use of census categories
25	Ethnic Monitoring Tool	NHS National Services Scotland/ Health Scotland	UK	All	Toolkit	The tool has been designed for NHS Scotland staff involved with the collection or use of ethnicity data	Explains need for data monitoring, who should be involved, what needs to be in place and provides some training materials
26	Who, when, and how: The current state of race, ethnicity, and primary language data collection in hospitals, 2004	Health Research and Educational Trust , The commonwealth fund	USA	All	Report	Survey and site visits to hospitals nationwide and report current practice, and identify problems	Reports inconsistencies in methods of collection, questions asked, and response categories. Report makes 5 recommendations: Standardise method of collection (self-identification should be used whenever possible) Point of data collection (admission ) Standardise categories, ideally US census Data storage should be standardised e.g. race and ethnicity stored as two separate variables Response to patient concerns and explanations should also be standardised
27	Gardi, M. Ethnicity data protocols training presentation, 2003	Ministry of Health, Manatu Hauora	New Zealand	All	Training presentation	Ethnicity data protocols, how to collect, classify, use ethnicity data	Ensure ethnic groups of policy importance are not swamped by NZ ethnic group. Each patient only appears once so sum of the population adds up to NZ population

28	Ethnicity: A review of data collection and dissemination, 2003	Social and Housing Statistics Section, Demographic and Social Statistics Branch, United Nations Statistics Division	UN	All	Report	Analysis of census data for countries including an ethnicity question. Report describes the ethnicity questions and responses allowed	The results show 107 questions were asked by 95 countries. These can be placed in five categories: 43% of questions used some form of categories with an open ended box for 'other', 20% had categories only, 21% were open ended questions, 4% had yes or no responses, 12% did not give enough information.
29	Ethnic group statistics: A guide for the collection and classification of ethnicity data, 2003	A National Statistics publication	UK	All	Guidelines	To suggest standards to ensure comparability of ethnicity data over time and meet the users needs	2 methods are proposed, one question (ethnicity) and 2 question (ethnicity and nationality). 2 question method should be used whenever possible
30	Diversity counts: Ethnic health intelligence in London, the story so far, 2003	London Health Observatory	UK	All	Report	Ethnicity monitoring issues in the NHS in London	Valid ethnicity data ranged from 17% to 100% by London's healthcare providers. Primary care identified as the poorest area, routine systems/integrated patient record could be possible solutions
31	Ethnic monitoring: A guide for public authorities, 2002	Commission for Racial Equality	UK	All	Guidelines	Ethnic data collection and monitoring guidance for employment, service providers, schools etc	Highlights the need for well designed mechanisms for ethnicity data collection and monitoring from dedicated personnel to databases and use of the data. Suggest method of collection also be recorded.

32	Collecting ethnic category data: Guidance and training material for implementation of the new ethnic categories, 2001	Department of Health	UK	All	Guidelines	Guidance for NHS staff collecting ethnicity data using the new 2001 categories and barriers to collection	Points explained include the new 16+1 codes, training for staff, and the importance of self-identification. There are brief summaries defining ethnicities and the usefulness of the data at a local and national level.
33	New federal standards for racial and ethnic data collection and reporting, 1998	Air Alert	USA	All	Guidelines	Changes to data collection following revised Office of Management and Budget (OMB) standards	Ethnicity data collection legal requirement for all federal agencies. Self-identification should be used wherever possible. Propose a 2 question method for self-reports and single question method for collection by observation
34	Patient profiling, 2005	Central Liverpool PCT	UK	All	Training presentation	Patient profiling in primary care	Important to collect through primary care as can use 'live' information for service provision.
35	HES online, 2004	NHS	UK	All	Report	Ethnicity coding in HES:1997-98 to 2002-2003	Overall records with missing ethnic data has decreased in the latest 5 year period

### 5.1.3 Summary of literature review

In summary, the main reason for rejection of a paper or article was the lack of description of the methodology of ethnicity data collection or ethnicity profiling. The systematic review identified 2404 papers of which 19 (0.8%) were deemed to contain information about methodology of data collection. In addition, there was a large volume of grey literature but upon review only 16 reports were included as relevant to this review.

## 5.2 Health Professionals questionnaire

There were 30 responders to the health professionals' questionnaire. The sample was well distributed throughout England & Wales, with 8 from Midlands, 6 from Wales, 8 from the North, 2 from the South and 6 non-stated. Breakdown of respondent by position in the NHS institution is shown in table 5.10, with the majority completed by clinicians, nurses or information officers. Role was not stated for 6 respondents.

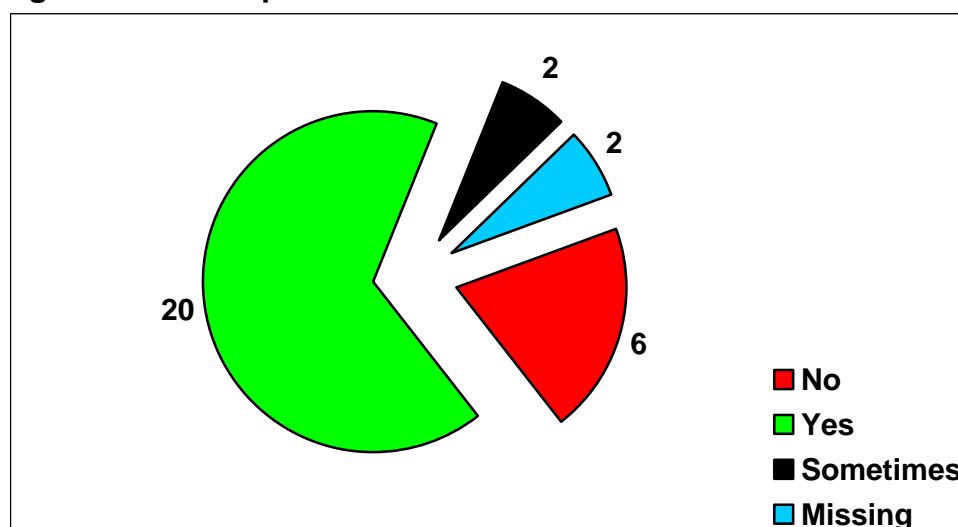
**Table 5.10: Questionnaire respondents by role**

Role	N
Clinician	7
Nurse	6
Information Officer	5
Radiographer	2
Cancer Services coordinator	1
Patient profiling development officer	1
Lead quality coordinator	1
Diabetes educator	1
Not stated	6
<b>Total</b>	<b>30</b>

### 5.2.1 Do you attempt to collect any ethnicity data on patients?

Two-thirds of respondents (n=20, 67%) attempted to collect some form of ethnicity data; 2 (6.5%) did not consistently collect any data, 6 (20%) did not collect any ethnicity data and 2 (6.5%) did not answer, see figure 5.1.

**Figure 5.1: Attempts to collect data**



## 5.2.2 Reasons given why ethnicity data are not collected

Of the 30 responders, 20 routinely collected ethnicity and 10 did not. The reasons given for not collecting ethnicity are stated in table 5.11.

**Table 5.11: Reasons ethnicity data not collected\***

<ul style="list-style-type: none"> <li>• “It is very difficult to record ethnicity data for our cancer records as it is not documented in the patient's case notes, to the best of my knowledge. Due to this, it would take a great deal of time to collect and is however, not asked for in any reports that are asked of me” (Cancer professional, Wales)</li> <li>• “Ethnicity data collection currently limited to Midwifery as Trust is taking part in the Welsh Assembly Government Patient Equality Monitoring Project and staff are awaiting training in how to collect information” (Human Resources Manager, Wales)</li> <li>• “We have not to date regarded it as sufficiently important” (Consultant, Wales)</li> <li>• “Not relevant to care or treatment given to patients. York has very few ethnic groups therefore language diet etc not required. Would access if appropriate” (Research Nurse, York)</li> <li>• “Only if it is required as part of a research trial and the company require that information. We then only fill it in, but it is very rare. We do not routinely collect this” (Research Nurse, Sheffield)</li> <li>• “Carried out retrospective 5 year audit to see if ethnicity influenced presentation with cancer, routes of referral, treatment received etc. Found study very difficult as ethnicity often not recorded on computer, had to check written notes.” (Consultant, Birmingham)</li> <li>• “Ethnicity data is not collected if it is not relevant. For example if an audit is being done and the question to be answered does not include an ethnicity component. Ethnicity data is difficult to collect because it involves asking the patient what they want it to be and they are not always available or willing to answer.” (Informatics Lead, London)</li> <li>• “Ethnicity data is not part of the datasets that are collected” (Information Manager, Yorkshire)</li> <li>• “Not part of my job” (Radiographer, Gloucestershire)</li> <li>• “Sometime ethnicity data is collected in the front of medical notes, but I expect the clerical staff don't understand the purpose of collecting such data” (Radiographer, Brighton and Sussex)</li> <li>• “Our data collection is poorly resourced as it is so we have to stay entirely focused on what is clinically relevant” (Oncologist, Birmingham)</li> </ul>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

\*Note: Responders identity presented as anonymous as possible without losing information



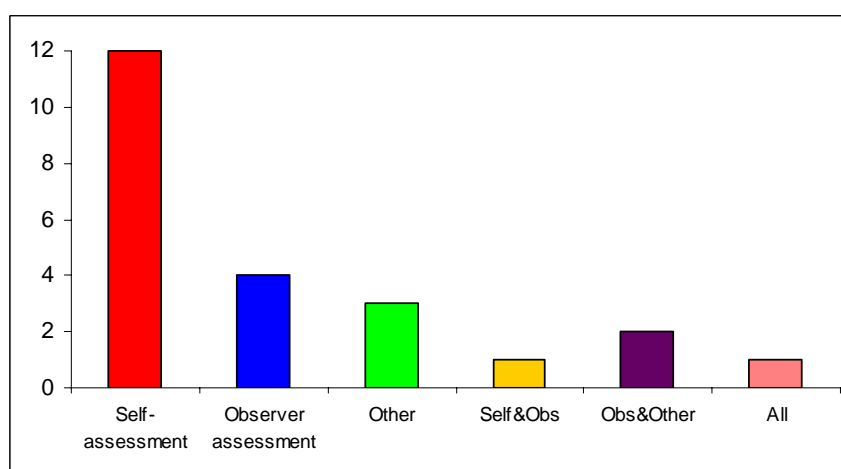
### 5.2.3 Disease areas for which ethnicity data are routinely collected

Of the 22 respondents who collected ethnicity data (routinely or sometimes), 37% collected it for cancer, 32% collected for all disease areas and 5% collected it for diabetes and hypertension. 'Other' areas included midwifery, all hospital registrations and contraception and sexual health.

### 5.2.4 Method of collection

The majority of respondents who collected ethnicity data used the recommended self-assessment method (n=12), observer assessment was used less frequently (n=4). Several respondents reported using a combination of methods e.g. self and observer assessment (figure 5.2).

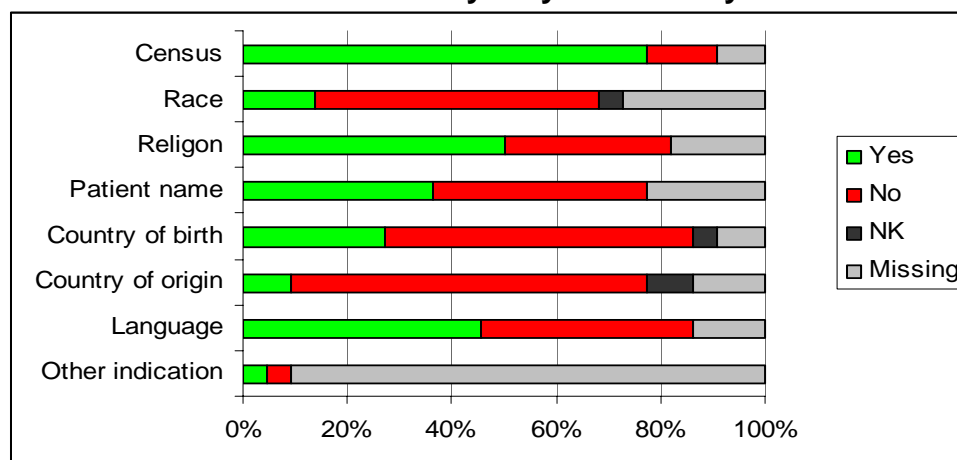
**Figure 5.2: Data collection methods (n=22)**



### 5.2.5 Indicators routinely collected

Ethnicity was most commonly collected using the census categories, other routinely collected indicators included religion followed closely by language. Country of origin, race, and country of birth were least likely to be collected. The 'need for an interpreter' was listed an additional item in 'other' (figure 5.3).

**Figure 5.3: which indicators of ethnicity do you routinely collect?**



Note: Other indicator was the need for interpreter

### 5.2.6 Problems encountered during ethnicity data collection

In the process of collecting ethnicity data respondents experienced a variety of problems, see table 5.12.

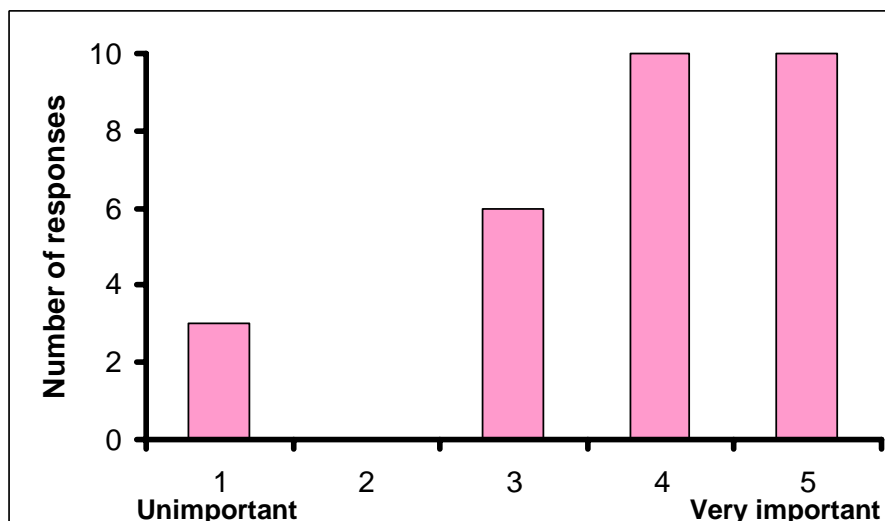
**Table 5.12: Problems encountered collecting ethnicity data\***

- “Collected as part of a large dataset & some items are poorly returned” (Cancer professional, Wales)
- “We depend on third parties in hospital trusts-poor data quality” (Welsh Cancer Intelligence and surveillance unit)
- “We have been collecting data surrounding ethnicity etc for around 7 years. The main issue is the patients’ lack of understanding of what ethnicity is. Also practice staffs lack of awareness of why we need to collect this information. On the whole though there have been very few problems.” (Patient profiling development officer, Liverpool)
- “We currently only record ethnic group in its widest sense” (Manager, Sandwell)
- “Often not recorded on software, so had to retrieve old notes and read through pages of clerking notes. Ethnicity usually recorded by junior doctors + written in. I did not wish to assume ethnicity from name alone.” (Consultant, Birmingham)
- “People collecting the data may not realise that they have to ask the patient.” (Informatics Lead, London)
- “Pt will ask why you need to know. If they come for anonymous info do not want to be listed. Do not accept that you need to have an idea of Ethnic origin so as to be able to review/develop/change service that is provided.” (Information & Support Services Manager, Birmingham)
- “I feel this is a difficult area due to fear of offending anyone. Most of the younger generation are British, I would have thought.” (Nurse, Birmingham)
- “Clients have the option of not stating their ethnic origin so there will always be a gap in the data” (Service Development Officer, Sheffield)
- “Failure of required process (i.e. patient not asked to self-select)” (Information Services, Bradford)
- “Vague 'Asian' (and similar for other groups) labels do not provide information due to heterogeneity of many groups” (Macmillan Cancer Information Facilitator)
- “Existing data collection systems are not made for it. Ethnic categories are not up to date, follow old traditional immigration routes” (Information Analyst, Luton)
- “We have had difficulty releasing the vast numbers of staff required to attend 'patient equality monitoring' training sessions, however this has been made easier by an all Wales Patient Equality Monitoring project sponsored by the Welsh Assembly Government and run by the NHS Wales Centre for Equality & Human Rights, who have produced an excellent Train the Trainer pack for Patient Equality Monitoring.” (Manager, Wales)

\*Note: Responders identity presented as anonymous as possible without losing information

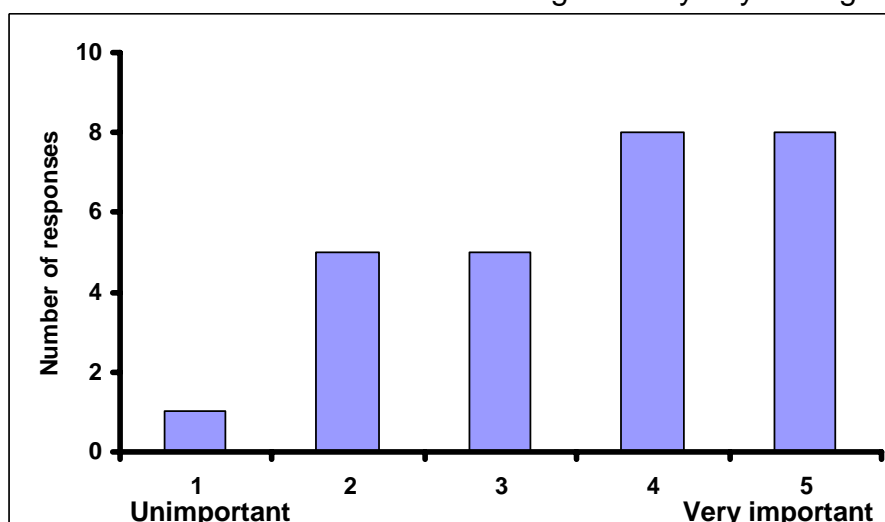
### 5.2.7 Perceived importance of ethnicity data collection

Figure 5.4: Please rate how important you personally think the collection of ethnicity data is?



The majority of respondents personally thought collection of ethnicity data was more important than their perception of its value to their organisation. This may be indicative of the organisations weak policies on ethnicity data collection and lack of training provision (figure 5.4 and 5.5).

Figure 5.5: Please rate the value of collecting ethnicity to your organisation



### 5.2.8 Summary

In summary, two-thirds of responders routinely collect ethnicity data with the commonest form of collection being self-assessment. Reasons for not collecting ethnicity stem from lack of understanding, lack of resources and lack of training. Of those who do collect the data, most perceive it to be important, and surprisingly a few of these perceive it to be less important to the organisation.

## 5.3 Patients' experiences and perceptions

Five focus groups were conducted by trained facilitators, each speaking in the preferred language of their group and also in English, if required. The number of participants in each group ranged from five to ten. The characteristics of the total 36 volunteer purposeful sample are shown in table 5.13.

**Table 5.13: Characteristics of participants**

Group	Country of origin	Language	Gender		Median age (range)	Total
			M	F		
1	Azad Kashmir	Mirpuri	0	5	-	5
2	Bangladesh	Sylheti/Bengali	8	0	63 (45-70)	8
3	Pakistan	Urdu	0	10	28.5 (18-35)	10
4	Pakistan	Urdu	8	0	30 (24-44)	8
5	India	Punjabi	2	3	31 (26-51)	5
<b>Total</b>			<b>18</b>	<b>18</b>	<b>31.5 (18-70)</b>	<b>36</b>

The facilitators used the topic guide (appendix 4) which was specifically developed to focus on the five areas of interest, together with sub-categories where appropriate.

### 5.3.1 General opinions on the collection of ethnicity

In general, participants thought accurate recording of ethnicity data was important. The majority were proud of their origins and were familiar with the differences between their culture and other cultures, understanding the potential usefulness of such data in a healthcare setting.

- “Everybody belongs to an ethnic group. Every single person. White, Welsh, Irish, Scottish, everyone” [Bengali male, participant 1]
- “Cancer Research UK should put more energy in alleviating the postcode lottery and make sure that treatment of all types of cancer is available to everyone irrespective of place where he lives and this is more important than ethnic data collection” [Urdu male]
- “Ethnic record keeping is important to determine how many different types of ethnic groups are in this country”. [Urdu male]
- “What are the problems of these ethnic groups”. [Urdu male]
- “The government want to know how many ethnic groups are living in this country apart from normal British, Scottish and Irish people”. [Urdu male]
- “Ethnic record keeping is more political than social”. [Urdu male]
- “England has a multicultural society and it is important to know ethnicity”. [Urdu male]
- “To establish everyone’s right in this country, it is important to carry out research. They are studying foreigners [immigrant people’s] culture only” [Bengali male, participant 4]

Several also knew the increased prevalence of certain diseases in minority ethnic groups and stated this as the reasons why ethnicity data collection is important in a healthcare setting.

- “Sometimes it is helpful to provide ethnicity as it helps care providers understand our background and determine common illnesses due to dietary habits or genetic findings, i.e. there are some health problems for which the incidence is far higher in Indians than is say for examples British whites so in these cases it would be useful to collect ethnicity. However, we should be told why it is being collected when asked for it” [Punjabi female, participant 4]
- “Because the service provider would have already done their research on the cultural differences and therefore the data would help to understand ethnic background much better and determine incidence of certain diseases” [Punjabi female, participant 5]
- “Sometimes certain illnesses are directly linked to our ethnicity. If a doctor does not know the right ethnicity he can not do any thing. For example stroke or diabetes is directly linked to ethnicity” [Urdu female]
- “Crucially diet is important. Not all of us eat the right type of food. Food varies from culture to culture” [Mirpuri female, participant 4]
- “It helps to understand culture or background, some illnesses are very genetic and higher in certain ethnicities” [Mirpuri female, participant 5]
- “Funnily enough I tend to see that GPs use it to find out your family background and relationships! But on a serious note I think it is to find out matters such as dietary lifestyle, i.e. high salt intake linked to increased heart attacks especially in Indians” [Punjabi female, participant 1]
- “To determine illnesses related to certain diets but also for research” [Punjabi male, participant 2]
- “So far as I know, they collect the data for different communities like Bangladeshis, Pakistanis, to carry out research with this data. This is called medical research; say you have diabetes, they want to know how many Bangladeshis suffer from diabetes, why they suffer from diabetes; how many Pakistanis, how many Somalis. Later they total up these figures to obtain another figure – the percentage for South East Asians altogether. People in South East Asia have been identified. ...” [Bengali male, participant 4]
- “They want to see from which country, which people suffer from which illnesses; where these illnesses come from and what is the cause of these illnesses? is it because of the weather in this country [U.K.] or their other country; they decide this. They can then understand that people of whatever country suffer most from whatever illness” [Bengali male. Participant 3]
- “In case of pregnancy, it is possible that the baby is carrying some disease in its genes this is also the reason the information should be provided” [Urdu female]
- “To certain of us it seems that ethnic data is irrelevant but to researchers it is important as they can draw links between different diseases” [Urdu male]
- “This also informs the government about the type of illnesses present among different ethnic groups”. [Urdu male]
- “As we know whites and Asian have some difference in the incidence of diseases and government wants to record ethnicity to observe and predict future trends in diseases”. [Urdu male]
- “It is used for statistical purposes, how many ethnic groups are out there and what kind of illnesses is common among these groups”. [Urdu male]

Several mentioned the importance of monitoring access and uptake of services. Others mentioned the need for collection of ethnicity for future planning:

- “So we know what kinds of people use which service or have what problem [Mirpuri female, participant 3]
- “If any research suggested, for example, that certain services were used by 20% of Pakistanis, 15% Somalians, 25% Indians; whereas less than 1% of Bangladeshis used the services: Would they [service providers] use this finding to encourage Bangladeshi to come forward to use the services [Bengali males, participant 7]
- “Future allocation of funds, resources and to determine what type of interest the government or private sector show “. [Urdu male]
- “This is important for future anticipation about health of ethnic minorities”. [Urdu male]
- “It is used for future planning”. [Urdu male]

Others felt that it was OK to provide ethnicity for health purposes but not for other reasons:

- “It could be alright with diseases but when you have to give this information while applying a job it would be felt like discrimination. In case of jobs the ability should be taken into the account instead of appearance or colour. In case of health, it could be OK but in case of jobs it is not right” [Urdu female]
- “It differs according to situation like if we are going for health service then it is acceptable as we are also getting some services in return but I don’t see any point of providing information for employment purposes” [Urdu male]

A small proportion (4 out of 36) did not understand the need for ethnicity data collection as they did not think it was relevant to treatment, or they felt they may be discriminated against if ethnicity is given:

- “Because ethnicity should never be a deterrent or an incitement when it comes to service or health provision so there’s no reason for why it should be collected” [Mirpuri female, participant 1]
- “Because we are all human and the same and so our ethnic origin should not interfere with the care we receive. After all it is our health that should be the main concern here. [Punjabi female, participant 1]
- “It is important for government point of view but there is no importance from our point of view [Urdu male]
- “This data can be used from the criminal point of view but there is no other use” [Urdu male]

### **Any objections/worries about providing this information?**

When asked about any objections or worries about providing ethnicity data the majority had no objections. Most of the participants' experience of providing ethnicity information was for job applications. In general this was not perceived to be useful information and many felt that it may discriminate against them getting the job. Several had concerns, or sometimes felt unease, if the purpose of data collection was not fully explained:

- “No objections as such but when I do provide this information on a job application and I don't get the job or if I provide the information to service providers and I don't get the right service then I cant help and think it was because of that” [Mirpuri female, participant 1]
- “It is best if it is done fairly without discrimination” [Mirpuri female, participant 4]
- “Sometimes patients may not be treated as individuals, we may judge by ethnicity and assume they have this problem as its high in their group” [Mirpuri female, participant 5]
- “Initially no problems at all but on a second thought I think it is not good when you provide ethnicity data and the prevalence of certain illnesses are automatically related to you” [Punjabi female, participant 1]
- “My only problem is when the category is not available on a form, e.g. British Asian, I very rarely see this category. However, I have no problems as the information is confidential and most of the time nothing is done with information apart from stored on their files for years to come” [Punjabi female, participant 4]
- “The NHS is so large with multi-cultural staff that I am not concerned I will be discriminated if my ethnicity is collected. However, I feel they should tell us when the information is collected and what it will be used for” [Punjabi female, participant 5]
- “I feel uneasy sometimes and you start wondering why they ask me questions about my ethnicity” [Urdu male]
- “No one tells us why they are asking such information” [Urdu male]
- “Sometimes one feels that they may use this data to discriminate me” [Urdu male]

### 5.3.2 Experiences of providing ethnicity information

In general when asked about how they felt, in their experience, about providing information about their ethnicity, the majority of people did not mind. Others only minded when they were asked at each repeat visit. The majority wanted some explanation as to why the data was being collected and what use it would be. The main reason given for negative experiences was inadequate codes for their ethnicity and the fact that on several forms they would be coded as 'other'. Positive experiences included one participant who did not speak English and was given a translator to help. All participants had no objections to providing ethnicity information in a healthcare setting. There was a strong feeling that the NHS is multi-cultural and providing this data would not lead to discrimination.

#### **Generally in any situation, how do you feel about providing information about your ethnicity?**

- “Not preferred“ [Mirpuri female, participant 1]
- “If someone asks then I will say. They must be asking for a reason“ [Mirpuri female, participant 2]
- “I don’t mind. I’m not ashamed of it so why should I mind“ [Mirpuri female, participant 3]
- “As I have mentioned as long as it does not discriminate towards others“ [Mirpuri female, participant 4]
- “I think it’s a good idea for patients to be able to be recognised, so they can be individually treated without a standard treatment for all“ [Mirpuri female, participant 5]
- “No problems“ [Punjabi female, participant 1]
- “No concerns but it should be stated why this information is being collected“. [Punjabi male, participant 2]
- “I just complete the form without giving much thought as to why I am providing this information“. [Punjabi male, participant 3]
- “Well, generally when you are given a form to complete you just fill it in and don’t question it whether it is for my benefit or not“. [Punjabi female, participant 4]
- “No problems“ [Punjabi female, participant 5]
- “My child was born in the same hospital yet they ask ethnic data about him whenever I took him to hospital“ [Urdu male]
- “Sometimes they ask these questions about ethnicity and sometimes they do not so we are not sure what is the standard routine“ [Urdu male]
- “No one tells us why are they asking such questions and I feel they should tell me why do they need this information“ [Urdu male]

#### **Have you ever had a negative or a positive experience of providing such information?**

- “Well I think I have like I mentioned before but you know with race legislation or so on, if someone is going to discriminate against you because of your ethnicity they



are not going to mention that that was the reason but often it is and its just covert“  
[Mirpuri female, participant 1]

- “No“ [Mirpuri female, participant 2]
- “No. You provide the information and it gets stored and that is it“ [Mirpuri female, participant 3]
- “I have never been in a situation that I can think of. It’s not about being negative or positive, but about being appreciated and recognised by the health service” [Mirpuri female, participant 4]
- “Most forms did not differentiate Asians, as Asian can be different groups, and not just Pakistani, not just Chinese, also people are living in Kashmir part of Pakistan do not like calling themselves Asian Pakistani, but want to be grouped as Asian Kashmiri, and recently that has been acknowledged” [Mirpuri female, participant 5]
- “I’ve not really thought about whether it has been a good or bad thing to give my ethnicity data. However, I do find that Asians don’t treat you as nicely as white people do- maybe it’s an insecurity” [Punjabi female, participant 1]
- “I once went for a check-up and was asked my ethnicity. As I didn’t understand English very well I was given a translator who was very helpful indeed and was able to understand my lifestyle and background quite easily. In this case I was grateful for my ethnicity being asked” [Punjabi male, participant 2]
- “When I have to state ‘Other’ as my ethnicity is not on the form and I feel even now my origin is not widely recognised” [Punjabi male, participant 3]
- “Yes when my category is not available and I have to state my ethnicity under ‘Other’”. [Punjabi female, participant 4]
- “I also find that as soon as a carer knows of my ethnicity I am treated better by White carers compared to Asian carers. Maybe Asians feel embarrassed about themselves? ” [Punjabi female, participant 5]

### **How do you feel about providing information about your ethnicity in healthcare situations?**

- “To ask asylum seekers or refugees questions about ethnicity or nationality I can understand because you know healthcare providers need to establish if the patient is an NHS patient or not, but for the general public I don’t understand what the significance is in healthcare situations. It shouldn’t be a factor so why is it important” [Mirpuri female, participant 1]
- “I normally take my kids with me if I have a hospital appointment so they fill out all the forms for me. I don’t know what they put” [Mirpuri female, participant 2]
- “I don’t mind” [Mirpuri female, participant 3]
- “Not embarrassed at all. We should be ashamed as to what we are. At the end of the day, we are all equal” [Mirpuri female, participant 4]
- “It shouldn’t matter” [Mirpuri female, participant 5]
- “I sometimes feel nervous that I won’t be able to understand what is going to be said to me” [Punjabi female, participant 1]
- “I feel fine about it as it concerns my health and I am usually open about not being able to speak English fluently” [Punjabi male, participant 2]
- “I feel fine about it” [Punjabi male, participant 3]
- “Generally, if you are born in the UK, you can speak fluent English and therefore get by but I do feel that regardless of giving my ethnicity we are stereotyped just by

skin colour and put into a category of “Asians”. What people need to realise is that there are many different types of Asian ethnic origins” [Punjabi female, participant 4]

- “Well, if we go to hospital they do ask us; I was asked this question and was happy to provide the information” [Bengali male, participant 2]
- “If you go to any clinic they ask certain questions, such as whether you have any disease, if you have cancer, breast cancer, arthritis, or heart problems, you have to complete this form, I did this many times” [Bengali male, participant 4]

**Do you think there are any effects of providing such information in healthcare situations? (Please give details of any experiences you may have had**

- “Yes, I don’t have any such experience myself but when my brother was admitted to hospital, before he was even allocated a bed or seen by a doctor we had to provide his demographic details. Makes me wonder whether based on that information he was given a bed next to the toilets or was that coincidence” [Mirpuri female, participant 1]
- “I don’t know but I don’t think so” [Mirpuri female, participant 2]
- “No” [Mirpuri female, participant 3]
- “None of the sort” [Mirpuri female, participant 4]
- “No” [Mirpuri female, participant 5]
- “I don’t know but I don’t think so” [Punjabi female, participant 1]
- “No” [Punjabi male, participant 2]
- “No because within the NHS everyone is works as part of a professional team and staff training involves strict discrimination laws. Also, patient’s needs are the same. It could also be that until you haven’t compared your experiences you may not realise the effects at all” [Punjabi male, participant 3]
- “There is such a multi-cultural society employed within the NHS that ethnicity is not an issue for patients. Also, NHS staff tends to look at the ethnicity information provided on the form but rather at your skin colour” [Punjabi female, participant 4]
- “The NHS deals with thousands of patients a day from all different cultural backgrounds that I don’t think they have time to discriminate. It may seem that we are getting discriminated by some staff but then it could be that an individual’s personality is not the best! “ [Punjabi female, participant 5]

### 5.3.3 Categories used in practice

When discussion was focused on categories used in practice to describe individuals, many participants wanted country of birth, language and religion to be collected, in order to be able to distinguish between 'South Asians'. One participant thought that additional information on diet was useful; another participant also thought it would be helpful to be asked whether you wanted to be a donor or not.

#### **What categories would you like with regards to providing personal information and– how would you prefer to describe yourself?**

- "I don't think there is a need for it at all. I would prefer to describe myself in the form of my date of birth, name and address and that's it" [Mirpuri female, participant 1]
- "I don't know because I can't read or write so I don't know and if someone is asking my children will answer for me". [Mirpuri female, participant 2]
- "The usual categories are ok. So Asian or Pakistani and Muslim and age and gender". [Mirpuri female, participant 3]
- "To represent a person who is not able to speak the language like in a Surgery as I have experienced". [Mirpuri female, participant 4]
- "Kashmiri" [Mirpuri female, participant 5]
- "I think we should be asked about Donor cards, as I find this information would be helpful" [Punjabi female, participant 1]
- "The current ones are fine but language would be good as there are cultural differences depending on what language you speak" [Punjabi male, participant 2]
- "If the information is for research then I think questions about dietary lifestyles should be asked. This is especially important as the younger generation tend to eat less Indian/Punjabi foods and more Western varieties. People should not assume that because we are Indian we always eat Indian food" [Punjabi male, participant 3]
- "My background is I am from Bangladesh, so British Bangladeshi, this is fine. My son was born and brought up here, so he will say British - that's it" [Bengali male, participant 6]
- "British Bangladeshi gives them accurate information for research [this was supported by two more participants]. For political reasons I say 'British Muslim', When it comes for ethnicity for medical research I would say British Bangladeshi" [Bengali male, participant 7, most of the others in the group agreed with him]
- "The more information you give the better it is. It's better to describe yourself British/Indian Asian" [Urdu female]
- "It is OK when there is a large minority ethnic group like British Pakistanis but what about when there is a very small minority group?" [Urdu female]
- "The ethnicity should not be confused with the colour of the skin" [Urdu female]
- "It is alright to mention the country of origin but if someone is born in the UK that fact should also be mentioned in the same form as well" [Urdu female]

### 5.3.4 Language, religion and culture

Overall everyone was happy to disclose their religion and language as long as they are not stereotyped. The discussion on culture centred on religion being a better indicator of culture than 'ethnic group'. Some Muslims did feel that they were stereotyped, especially with new awareness of terrorism. Also some Bengalis have been called 'Paki' based on their skin colour, which is stereotyping. Others did not feel stereotyping to be a problem and are proud of their language, religion and culture and do not mind providing this information.

#### ***Do you get asked, do you offer this information, and do you mind?***

- "I have been asked, I have provided only because I'm not ashamed of my religion and whether I mind would depend on why I'm being asked" [Mirpuri female, participant 1]
- "My children will answer. They will say" [Mirpuri female, participant 2]
- "Yes I have given this information before and I don't mind" [Mirpuri female, participant 3]
- "As long as it doesn't get out of control!" [Mirpuri female, participant 4]
- "No I don't mind, as long as I am not stereotyped or characterised I should still remain an individual as I am" [Mirpuri female, participant 5]
- "I do get asked but I don't mind" [Punjabi female, participant 1]
- "Yes, but it's perfectly fine" [Punjabi male, participant 2]
- "If I do I just ask them back why this information is necessary" [Punjabi male, participant 3]
- "Yes but I think this is just to make the patient feel comfortable. However, there is the case of Jehovah's Witnesses where it is necessary to give information about Religion so I guess there are exceptions" [Punjabi female, participant 4]
- "Yes but usually for medical reasons so I don't mind" [Punjabi female, participant 5]
- "I would not hesitate to describe my language as Bengali, no reason to feel *"sonkuchito"* ["sense of shame"- others agreed with him] " [Bengali male, participant 3]
- "I am not clear about culture, the reason is that as a Muslim, my culture is Islamic; Islamic Culture and Bangladeshi culture are not the same thing; many time we see differences between Bangla culture and Islamic culture. So if say my culture is "Islamic" is there any problem?" [Bengali male, participant 7]
- "I have seen in a questionnaire, "Bengali culture" mentioned. In Bengali culture there are many things which are Hindu activities, and these relate to Hindu culture rather than Islamic culture. There is no relationship between this Hindu culture and Muslim culture or Bengali culture. These are not our culture. Therefore in many times what we call Bengali culture, in reality, if we examine further, we will see that these are against Islam. For example, the celebration of morning procession on first of "*Boyshak*" [1st day of Bengali calendar]. *Boyshaki mala* is not considered part of Bengali culture". [Bengali male, participant 4]
- "Religion should be a part of ethnicity because that is the base of one's lifestyle and dietary requirements. We do not know if the medicines we are taking are in accordance with the dietary requirements of our religion e.g., most of the cough medicines may have alcohol in them" [Urdu female]

- “Language is important because sometimes an interpreter may be required. The participants were confused on how much information should be asked e.g., should they be asked about the mother tongue? How much fluency one has in which language? British born people often get confused on it because English is their main language and they have got fluency in it but still they are asked questions about their language“ [Urdu female]
- “There is no problem in disclosing one’s religion but most of the people felt uncomfortable when disclosing one’s culture. Most of the British born people had very little experience of their native countries and so had very little idea of the culture their. So they do not know about their true culture except from the daily Asian meals. They consider themselves possessing more British culture than the culture of their ancestors and consider culture very vague“ [Urdu female]
- “Religion is more important as most of us do not compromise on religion“ [Urdu male]
- “Language is also important for those who cannot speak English“ [Urdu male]

***Do you think there is a risk of ‘stereotyping’***

- “Yes I do. Like I said before, it’s all covert so there’s no evidence of it but its the feeling that I get and my gut feelings are mostly correct“ [Mirpuri female, participant 5]
- “No“ [Mirpuri female, participant 5]
- “Maybe but I don’t think so“ [Mirpuri female, participant 5]
- “There is always that risk in everyday life, but I guess people are far too busy with other duties to take notice“ [Mirpuri female, participant 5]
- “Yes, if certain members of the community or culture do not agree or follow certain ways, it doesn’t mean everyone will be the same, choices, independence to decide, options and opinions should be offered and noted“ [Mirpuri female, participant 5]
- “Yes, I feel that I am regarded as a vulnerable women because I am a non-English speaking person“ [Punjabi female, participant 1]
- “No comment“ [Punjabi male, participant 2]
- “Walk-in centres provide independent advice but I feel my GP knows my family history so makes assumptions about me“ [Punjabi male, participant 3]
- “Well, at the GP surgery I don’t feel stereotyped as my GP knows more about me. However, when I visit the hospital my experience is different as I feel the staff quickly perceive you as having the same background to someone else who has the same skin colour as me“ [Punjabi female, participant 4]
- “Usually I feel stereotyped at the GP surgery but not in an NHS walk-in centre. I feel the GP makes assumptions about his patients“ [Punjabi female, participant 5]
- “I have had such an experience. One day someone called me “PAKI”! I am not a ‘Paki’, they all think we are “Pakistanis”, we are not, we Bangladeshis have different ways of thinking. This is stereotyping. They are colouring us all the same“ [Bengali male, participant 7]
- “I am not Pakistani, I am a Bangladeshi. Because of my colour and appearance someone is calling me “Paki”. This is stereotyping“ [Bengali male, participant 2]
- “About the stereotyping, it depends how you look at it: Muslim is OK“ [Bengali male, participant 7]

- “If some one says Muslims are terrorist, I am happy to be known as Muslim, Allah will help us in Heaven. We are true people. When it is said “terrorist” it alienates us“ [Bengali male, participant 4]
- “The suspicion is that all Muslims are terrorist. This is a stereotyped view. This kinds of stereotype views should not be allowed“ [Bengali male, participant 7]
- “If some one has a beard they look at him and consider him as terrorist, this is wrong“ [Bengali male, participant 2]
- “Fear of stereotyping is there. Any brown complexion person may be called a Paki or a girl with head scarf may be labelled a terrorist. This is the main fear of disclosing one’s origin“ [Urdu female]

### 5.3.5 How information should be collected?

The Bengali focus group summed up how information should be collected:

- “They should explain why they collect the data; the reason behind it; what benefit there will be for people. Also, where the data will be used and how secure this data will be. It should be kept secret [confidential]” [Bengali focus group; all participants].

Most participants agreed that GPs should collect ethnicity data once and that this should be available to hospitals. There was a general consensus that not enough information is provided as to the use and importance of this data.

#### ***Would you recommend the routine collection at hospital/GP/other?***

- “No way. There is no need for routine collection. If it really has to be it only needs to be collected once at each institution“ [Mirpuri female, participant 1]
- “If it’s needed“ [Mirpuri female, participant 2]
- “Not routine; there is no need since these things don’t change but once or twice is ok“ [Mirpuri female, participant 3]
- “Where ever convenient and suitable“ [Mirpuri female, participant 4]
- “Yes“ [Mirpuri female, participant 5]
- “The information should be collected at the GP surgery as patients are already distressed in hospital“ [Punjabi female, participant 1]
- “Yes definitely“ [Punjabi male, participant 2]
- “OK to collect it“ [Punjabi male, participant 3]
- “The information should be collected ideally as soon as you have registered with your GP“ [Punjabi female, participant 4]
- “Yes fine to collect it“ [Punjabi female, participant 5]

#### ***When would be the best time to collect this data (admission/follow-up after you’ve been to the hospital once)?***

- “I think to ask at admission would dishearten patients from the beginning so that wouldn’t be a good idea. Follow up would be better because then the patient would have had a chance to get to know their doctor and gain confidence“ [Mirpuri female, participant 1]

- “Whenever is fine with me. But admission would be the best I think” [Mirpuri female, participant 2]
- “Admission” [Mirpuri female, participant 3]
- “During appointment” [Mirpuri female, participant 4]
- “When registering at your GP surgery, which is your first port of call for medical purposes. A+E also collects ethnicity so it is registered in your medical notes at the hospital” [Mirpuri female, participant 5]
- “It should be the GP’s responsibility to provide the information” [Punjabi female, participant 1]
- “I also think it should be the GP who should give this information” [Punjabi male, participant 2]
- “If admitted as an emergency at the hospital then should be given at the hospital” [Punjabi male, participant 3]
- “If in hospital then once admitted and patient should be given the option to complete the form at their own leisure!” [Punjabi female, participant 4]
- “Hospitals should not take samples or have us filled in such forms every time we visit there because all the data information is online and they can access our data from the online database. We should inform them only when any of our circumstances have changed to update the online database” [Urdu female]
- “Some participants pointed out that about 10 to 20 years ago it was quite comfortable to communicate and give information and they were not required to give too many details but now things have changed and information requirements are becoming more and more complex” [Urdu female]
- “Cancer register is just starting in our ethnicity group” [Urdu female]
- “They should ask about our ethnic data once treatment is over especially during emergency situations” [Urdu male]

### **How should staff ask you – and what should they tell you?**

- “They should tell us why they need this information. A reason other than its just procedure. And who has access to the information” [Mirpuri female, participant 1]
- “I don’t speak English so they don’t ask me. I kids will normally say but if there is an Asian doctor or nurse then I can tell them” [Mirpuri female, participant 2]
- “Just generally ask or give you a sheet to fill in if you can read and write” [Mirpuri female, participant 3]
- “In writing as it is sensitive information and people can be easily offended if not said the correct way!” [Mirpuri female, participant 4]
- “Give a list of all options, read them out, allow them to choose correct option, and tell patients it for computer audit purposes, and not for patients personal data and that it will not affect the way they are treated for their condition” [Mirpuri female, participant 5]
- “Maybe by means of a letter which you can read at your own leisure and this can be followed by verbal explanation” [Punjabi female, participant 1]
- “Copies should be retained by the patient and the carer as a record of the information provided” [Punjabi male, participant 2]
- “Maybe provide information on how this data has helped previous cases” [Punjabi male, participant 3]

- “We should be told what is going to be done with this information“ [Punjabi female, participant 4]
- “We should be told why this information is being collected and it should also be made optional to provide it. We should sign a form to ensure that we have understood the information provided“ [Punjabi female, participant 5]
- “Written forms tell us why that kind of information is needed but no one explains us whenever they ask verbally“ [Urdu male]
- “If they explain its importance before they ask question about ethnicity, I will feel more comfortable in providing such information“ [Urdu male]

### 5.3.6 Overall summary

- In general, there was no objection to providing data for healthcare purposes
- Some understanding of differences in disease patterns, e.g. higher incidence of diabetes in South Asians, and the importance of recording these trends
- Staff should be able to explain why they need the information other than ‘procedure’
- Explaining why data is needed and what it will be used for would increase willingness to provide ethnicity data
- Ideally, ethnicity should only be collected once by GP or at first hospital visit
- There was a feeling that data collected for ‘statistical purposes’ is not fully utilised
- ‘Ethnicity’ information should include language, religion and country of birth to account for cultural differences



## 5.4 Validation of General Practice data

One General Practice Surgery in Nuneaton was visited to undertake validation (completeness and accuracy) of ethnicity data collected by the surgery. The surgery has 15,058 patients enrolled. Ethnicity has been recorded for all new patients enrolled since April 1, 2006. Of the 1480 newly enrolled patients 141 (9.5%) did not state their ethnicity.

To validate the Nam Pehchan software, it was run on the names and surnames of those who self-declared their ethnicity (gold standard definition). Out of 1339 patients, the Nam Pehchan classified 123 as South Asian and the rest 1216 as Non-Asian (see table 5.14). Its validation with the gold standard definition (i.e. GP record) showed a sensitivity of 67.5% and specificity of 97.1%. The positive and negative predictive values were 70.3% and 96.7% respectively.

**Table 5.14: Identification of ethnicity by Nam Pehchan and comparison with self-declared ethnicity**

Self-declared ethnicity	Nam Pehchan ethnicity		Total
	South Asian	Non-South Asian	
South Asian	83	35	118
Non-South Asian	40	1181	1221
<b>Total</b>	<b>123</b>	<b>1216</b>	<b>1339</b>
<b>Sensitivity=67.5%, Specificity=97.1%</b> <b>Positive predictive value=70.3%, Negative predictive value=96.7%</b>			

One of the authors of the report, who is also the name analyser, independently identified South Asian names from the list. He identified 117 names of South Asian origin. When compared with the gold standard definition, sensitivity worked out to be 75.4% and specificity of 97.4%. The positive and negative predictive values were 72.9% and 98.5% respectively (All these values found out to be greater than the previous results, table 5.15). When checking which names were left out and had self-reported as South Asian, we found out that all these persons were carrying Christian names and had migrated from the southern part of India (especially Kerala). This was also confirmed by other facts when we looked at their place of birth data.

**Table 5.15: Identification of Ethnicity by Manual Checking and Comparison with self-declared ethnicity**

Self-declared Ethnicity	Manual Checking		
	South Asian	Non-South Asian	Total
Not stated	13	128	141
South Asian	86	32	118
Non-South Asian	18	1203	1221
<b>Total</b>	<b>117</b>	<b>1363</b>	<b>1480</b>
<b>Sensitivity=75.4%, Specificity=97.4%,</b> <b>Positive predictive value =72.9%, Negative predictive value=98.5%</b>			

## 6 Synthesis of the results of the project

Identifying good evidence for the collection of ethnicity data and monitoring was the focus of this project. Reviewing the literature, health professional's experiences, focus groups and validation of existing databases identified a paucity of published evidence but a wealth of reports on the 'use' of ethnicity data. Of the 2404 possible published articles, only 19 (0.8%) remained in the systematic review; the main reason for rejection was description of 'use' (57%) but no mention of the methods for data collection. A systematic review of Ma (2007) also found a lack of reporting of the methods when examining reporting practice of race/ethnicity and socioeconomic status in biomedical journals [1].

Ethnic groups are different between and within countries and have to be defined on a national level. Reports from North America will concentrate on 'Hispanics/Latinos' who are their largest ethnic minority group as well as American Indians and Alaska Natives [2-4]. Other reports/guidelines such the New Zealand Ministry of Health indicate the need to ensure that ethnic groups of policy importance such as Maori have their own category and ethnicity is assigned using a priority system [5]. Within the UK South Asians are the largest ethnic minority group, accounting for 50% of the ethnic minorities in 2001 [6]. The meaning of the same ethnic group is also different across countries with 'Asian' referring to Chinese or South-East Asian origin in North America, compared to 'Asian' referring to 'South Asians' from the Indian sub-continent in the UK.

The ethnicity indicators collected are different between countries with the North Americans collecting ethnic group/race as standard [7]. Other indicators are language, religion and culture as demonstrated by a report by the United Nations [8] who identified 107 questions asked by 95 countries collecting ethnicity data as part of their census.

Ethnicity 'use' includes the selection of ethnic groups such as 'Chinese' or 'South Asians' for targeting health promotions or interventions. Many articles talked about multi-ethnic samples where ethnicity was not recorded but was the basis of a representative sample; e.g. smoking cessation in a multi-ethnic deprived area. These articles were rejected as they were not informative in terms of providing good evidence about ethnic data collection or monitoring. Other rejections were where articles talked about the population of their own country, e.g. multiple births in Saudi Arabian desert climate. Hence ethnic groups were a keyword in the Pubmed search criteria but such cases were obviously not included in this review. Better ways of tagging 'ethnicity' in Pubmed are needed.

Of the articles included in the review, 19 (54%) were identified from published literature and a further 16 (46%) from grey literature.

## 6.1 Ethnicity data collection and monitoring

Guidelines and examples of best practice were the most informative in determining methods for ethnicity data collection and monitoring:

- UK Commission for Racial Equality (2002)
- UK Department of Health practical guide to ethnic monitoring in the NHS and social care (2005)
- USA The Robert Wood Johnson Foundation Report (2006)
- UK Lambeth PCT review (2006)
- USA Health Research and Educational Trust (HRET) online toolkit (2007)
- NHS National Services Scotland Ethnic Monitoring Tool (2005)

These grey literature reports present the best practice evidence for coding of ethnicity, data collection, overcoming barriers to collection and incompleteness of data as well as interventions and best practice policy. None of these reports discussed the validation of data collection or use of alternative tools, such as use of the name recognition software for checking the data. However they all discuss the importance of collecting ethnicity, the need for training of staff and the need for complete data. The lack of discussion about validation is probably due to the consensus being that self-reported ethnicity is the gold standard and hence validation is not mentioned. Several did discuss the need for future verbal verification of the self-reported categories. Examples of best practice in the UK are given in the Department of Health report [9]. Key reports where ethnic data collection has been successful due to adequate resources, awareness and training [10, 11], also demonstrated the need to have a 'use' for the data in order to improve collection.

Site visits revealed that although there was a good level of commitment to collect the data, the absence of any standardisation made it difficult in practice [12]. This is reinforced by the focus group discussions where several participants found lack of information or lack of standardisation confusing.

Weinick et al (2007) describe the lessons learnt from implementing a publicly mandated racial and ethnic data collection program in Boston and Massachusetts acute care hospitals. The data collection tools must be standardised across hospitals, whilst continuing to meet the hospitals needs and accommodate the way in which patients chose to describe themselves. Staff training is essential when implementing new policies [13].

Ford et al (2005) carried out a critical review of published literature, providing an overview of the methods currently used to assess race and ethnicity in USA health services research. The importance of consistent and efficient methods of collecting ethnicity were discussed and the need to implement better data collection tools for more accurate reporting, especially in a Veterans Affairs population (the focus of the report) who are becoming increasingly diverse in race and ethnicity [14].

'Who, when, and how: The current state of race, ethnicity, and primary language data collection in hospitals' report [12] makes five recommendations for improving ethnic data collection and quality as follows:

1. Hospitals need to standardise the method of collection (self-identification should be used whenever possible)
2. Point of data collection e.g. on admission/registration recommended
3. Categories used for race and ethnicity should be the same across the board, ideally (US) census but refinement is supported as long as data can be combined into census categories
4. Data storage should be standardised e.g. race and ethnicity stored as two separate variables. New systems allow the data to be merged with clinical data records and be imported or exported
5. Patient concerns should be addressed prior to collection, response to concerns and explanations of data use should also be standardised.

The UK DoH has implemented policy change within the primary and secondary care settings. The impact of accurate ethnicity data collection has not been fully realised as there is still a long way to go before the data are complete and reliable [12, 15]. It is important to collect accurate ethnicity data for planning and policy making.

Categorisation, ethnicity indicators, methods of collection, barriers to collection, interventions, completeness of data quality and completeness collection will be discussed under separate headings.

## 6.2 Categories for defining ethnic group

Differences in ethnic groups between countries means that coding categories are different depending on which country the research is carried out. A United Nations report identified a total of 107 questions asked by 95 countries for collection of ethnic group within the census. Only 12% of countries collecting ethnicity data had categories for 'mixed' identities' or allowed multiple box selection [8]. Other international guidelines indicate that the gold standard categories used within a country may be expanded as long as they can be concatenated back to the standard categories, for national reporting purposes [10, 13, 16].

Also there are inconsistencies with the data types being used, with up to 5 difference types:

- Coded categories with text boxes
- Coded categories without text boxes
- Yes/No questions
- Free text self report
- Unknown response

The UK gold standard ethnicity categories are taken from the 'ethnicity' census question on 'ethnicity'. Ethnicity was first collected in the 1991 census where 10 categories were collected, this was further expanded in 2001 to the 16+1 categories which introduced a 'mixed' category and expanded the 'White' category (table 6.1). The CRE report and the Department of Health guide to ethnic monitoring both state the importance of not allowing patients to see the (+1) category of 'not stated' nor should the patient be verbally offered this option [9, 16]. The other difference between the 1991 and 2001 census categories was the ability to state ethnicity in more detail by a free text field if 'any other' was ticked across all ethnic groups.

**Table 6.1: 1991 and 2001 Census categories**

1991 Census – Ethnic Group		2001 Census – Ethnic Category	
0	White	<b>White</b>	
1	Black Caribbean	1	British
2	Black African	2	Irish
3	Black other.....	3	Any other white background.....
4	Indian	<b>Mixed</b>	
5	Pakistani	4	White and Black Caribbean
6	Bangladeshi	5	White and Black African
7	Chinese	6	White and Asian
8	Any other ethnic group.....	7	Any other mixed background.....
9	Not known/Not given	<b>Asian or British Asian</b>	
		8	Indian
		9	Pakistani
		10	Bangladeshi
		11	Any other Asian background.....
		<b>Black or Black British</b>	
		12	Caribbean
		13	African
		14	Any other Black background.....
		<b>Other Ethnic Groups</b>	
		15	Chinese
		16	Any other ethnic group.....
		<b>Not stated</b>	

### 6.3 Other Indicators of ethnicity

The UK DoH guidelines (2005) [9] encourage the collection of religion (Buddhist, Hindu, Jewish, Muslim, Sikh, any other religion, not stated), diet (No special requirements, vegetarian, vegan, restriction according to religion such as Kosher or Halal and food allergies or medical conditions), language and the need for an interpreter. These additional indicators of ethnicity should be collected only if they are relevant for a particular facility or local area, and analysed locally. In particular information about diet is helpful for in-patient stays and language to assess the need for an interpreter.

The Office of National Statistics (ONS) recommend that Nationality is also collected for planning and resource purposes (6 codes: English, Scottish, Irish, Welsh, British and Other) being the categories offered [17]. Responses should be re-ordered or altered depending on where the question is been asked, e.g. in Scotland Scottish should be at the top of the list. This reordering to put groups of policy importance is also practiced in other countries, such as New Zealand where 'Maoris' are at the top of the coding list [5].

The Individual Patient Registration Profile (IPRP) used in Lambeth PCTs collects self-reported ethnicity, religion, language and need for an interpreter in addition to the usual data items such as gender, date of birth, smoking status. Ethnic groups have been expanded for the local community but can be collapsed down to the census 2001 categories [10]. This is an ongoing 3 year project which attempts to collect demographic data, including ethnicity, on a central database (Datanet). The Datanet project is a resource which has been developed to link the IPRP data to research projects. Central Liverpool NHS Primary Care Trust have also carried out patient profiling by collecting detailed ethnicity data including spoken language and read language, as well as additional information about general health and carer information [18].

Indicators of ethnicity and their usefulness are discussed by Gill et al (2006) as part of the needs assessment series. Country of birth has been asked since 1841 but is no longer a reliable indicator of ethnic origin as 50% of minorities are born in the UK. Family origin or ancestry is a reliable indicator for individuals of single origin but problematic for mixed races. Name recognition software has proven to be a useful tool for identifying South Asians and can be used for validation purposes [19].

In summary, other indicators of ethnicity apart from ethnic group are country of birth, nationality, language spoken at home, country of origin in conjunction with country of birth, skin colour (white and black), national/geographic origin, diet and religion. The UK Department of Health guidelines (2005) give 17 examples of good practice throughout the UK with a variety of indicators collected in practice, demonstrating that these are collected depending on local needs [9].

The survey of health care professionals showed that the commonest indicators collected were the census ethnic group category, religion and language (figure 5.3). The need to collect other indicators as well as ethnic group was confirmed by the focus group participants who also wanted religion, diet and language to be collected; this is necessary for resource planning and service provision.

## 6.4 Methods of collection

Self-reporting of ethnicity is the gold standard and the reasons for this are discussed in many good practice guidelines and papers [9, 11, 16]. There is good evidence that ethnicity should be self-reported as health care professionals are not trained to determine ethnicity by observation. Observation leads to 'stereotyping' by skin colour and name and should only be used where self-reporting is not possible, e.g. language barriers. In the USA the paper by Hasnain-Wynia et al (2006), illustrates how staff should ask for this data and emphasises the need for self-reported ethnicity [20]. The HRET toolkit includes training on how data should be collected and again emphasises the need for self-reporting [21].

Results from the USA surveys conducted by the Robert Wood Johnson group revealed 61% of respondents said they usually asked the patient to self-identify, but 25% filled in the ethnicity themselves by observation. They felt this method was easier for them and the patients as it avoided any discomfort, they also felt it was accurate as they believed they knew their local population [11]. Self-reporting was the main method used by the health care professionals from the CanEth survey (figure 5.2) with 59% using this method of collection; but a substantial percentage do not use self-reporting.

Self-reporting has been described as a two step process, the first step is an individual's view of their own identity and the second the choice of responses. Ideally categories and the wording should be piloted prior to general use to ensure they are suitable and the language appropriate for the population [17].

Since self-reporting is important, the method of collection should be recorded, i.e. self-reporting, observation or other [16], and important biases could occur if assumptions are made about the reporting. Sugarman et al demonstrated that racial disparity varied according to method of collection and the incidence of renal disease in American Indians/Alaska Natives increased from 268 per million to 312 per million after corrections to coding [2].

In the UK no restrictions are imposed on the number of boxes an individual can select [9, 17]. This is also the case in New Zealand where their training kit recommended that participants should tick as many boxes as they want [5]. However, this is not the case internationally where only a small proportion of countries allow the selection of multiple boxes or have specific categories for mixed race individuals [8].

There is evidence to show that self-assessment of ethnicity is the most effective way to obtain accurate information, especially for the 'mixed' or 'other' ethnic groups. Buescher et al (2005) found discrepancies between published data on racial classification and self-reported race. In the study of 118,000 live births in North Carolina in 2002, mothers reported over 600 different versions of race on birth certificates and 2/3rds of mothers from Hispanic origin would have been coded as 'Other' from official statistical codes [22]. This shows the need to re-examine official categories, especially as the ethnic and racial diversity continues to increase. This frustration with official categories was also found in

the focus group participants, some of which requested that their small ethnicity group is collected, e.g. Kashmiri.

In addition, it is important to distinguish between the 'not stated' and occasions when staff 'fail to ask', as they are a different set of problems [9].

Other methods of collection could include the use of name recognition by observation or special software. This is not ideal, especially with an increase of mixed marriages, but it could be a useful tool for the capture of missing data. Patients notes were successfully used to identify most patients used in one study [23] which demonstrated that names can be used with some precision when no other data is available. However, it has been shown that name recognition software used in conjunction with other indicators such as country of birth data results in increased accuracy [4, 24, 25].



## 6.5 Barriers to collection

The main barrier to ethnicity data collection is staff lack of knowledge about the importance of the data and its uses. This was a very clear message from the healthcare professionals' survey (table 5.11) where the main reason for not collecting ethnicity data was the perception that 'it was not important' and that 'it is not relevant to patient care'.

Site visits to 6 consortium member hospitals in the USA and a 1,000 nationwide hospital survey to identify current data collection practices, found that 30% of respondents reported problems/barriers to collecting these data [12]. The barriers reported were similar to those found in the Robert Wood Johnson Report [11]; these being:

1. Reluctance of staff to ask this type of information for fear of offending the patient or encountering resistance
2. Confusion about race/ethnicity categories
3. No demonstrated need to collect this data
4. Reluctance of patients to provide this type of information
5. Limitations of health information technology to capture this type of data
6. Lack of resources (time constraints during registration process, costs of collecting/monitoring data, cost of database and staff training) and lack of funding to support the collection of this data
7. Concerns that ethnicity data collection may expose the hospital to legal liability
8. Lack of agreement of executive leadership on the need to collect this data

Other barriers include data confidentiality issues and restriction of some categories [12].

Many reports state that there is a need to use the data for research projects, set targets and compare outcomes. "The research should be published, this will motivate people to collect and provide data as they will know that it is been used for meaningful research/changes/commissioning" [10]. Until data are collected and reported, then no-one will think these data is useful or important.

The other main barrier is the patients' perceptions. Patients would be more willing to provide data if the reasons why the data are being collected are explained, also healthcare professionals should be comfortable asking for these data [26]. This was a strong theme for the focus group participants who would willingly provide ethnicity data for healthcare purposes, but want to know how the data are used. Baker et al (2007) reported 46% of patients worried that the data would be used to discriminate against them [27]. The focus group felt that for healthcare purposes it was fine to collect the data but in job applications they felt there was a danger of discrimination.

## 6.6 Interventions

All of the best practice guidelines recommended that the main intervention required for completeness and accuracy of ethnicity data collection was staff training followed by adequate resources for data collection and 'use' of the data [9-11, 16, 21, 28].

The 2005 UK NHS 'practical guide to ethnic monitoring in the NHS and social care' states that staff training should explain why ethnic monitoring is important and what the data will be used for, how to collect it, and tailor training to the local population. It should also promote the self-assessment method as been the most accurate method of collecting this type of data. Local community groups could be asked to comment on the content of the training packs. All staff that may be involved in collecting ethnicity data, write reports or analyse data or make decisions based on the data, need to attend training. Training needs may differ from group to group. Self-classification is crucial as ethnicity is part of an individuals' identity that may/may not be seen by a third party [9].

In the USA, the HRET toolkit provides a national training package for the collection of ethnicity data (<http://www.hretdisparities.org/>) It is written for all levels of healthcare workers including Chief Executive Officers, Clinicians, registration staff, database managers as well as patients so that users can pick the training package/information most relevant to them. Once registered, the toolkit is free [21].

The content of the toolkit is as follows:

- Who should use the Toolkit
- Why collect race, ethnicity, and primary language data
- Why collect data using a uniform framework
- The nuts and bolts of data collection
- How to ask questions about race, ethnicity, and primary language
- How to use race, ethnicity, and primary language data to improve quality of care
- How to train staff to collect this information
- How to inform and engage the community
- How to address the communication access needs of deaf and hard of hearing populations
- Available tools and resources
- Answers to frequently asked questions

Resources available on the website include:

- Overview presentation on collecting race, ethnicity, and primary language data
- Staff training presentation on collecting race, ethnicity, and primary language data
- Presentation on addressing concerns from patients with applicable questions and answers
- Survey on collection of race and ethnicity data by hospitals

- Office of Management and Budget's race and ethnicity definitions
- Office of Management and Budget's granular code set on race and ethnicity
- Centers for Disease Control and Prevention's granular code set on race and ethnicity
- Reference booklet for staff on data collection categories

The training emphasises the importance of prioritising the individual's view of their own ethnicity and the need for self reporting.

Apart from the best practice guidelines in the UK, the most comprehensive training package is the 'Ethnic monitoring tool' developed by NHS Scotland (<http://www.isdscotland.org/isd/files/ETHNIC%20MONITORING%20TOOL.pdf>). This is aimed at NHS Scotland staff and provides information on why to carry out ethnic monitoring, who is involved in monitoring and what needs to be in put place. Training materials, including PowerPoint presentations, can be downloaded and modified according to local needs. Also train the trainer notes and role play scenarios are provided [28].

The Lambeth PCT project provides 1.5 days training (1 away and 0.5 in the practice) for staff to understand the need for ethnic monitoring, practicalities of data collection and data use. Computer templates are provided and resources available to mailshot a questionnaire to existing patients as well as collecting ethnicity for newly registered patients [10].

An additional training programme has been developed through the Cancer Research UK PROCEED project which is a multimedia tool for educators within the health and allied professions. PROCEED "helps with the development of skills for responding appropriately to patient diversity by raising awareness and understanding of cultural diversity" (<http://info.cancerresearchuk.org/proceed>). However it does not provide training for the collection of ethnicity data [29].

The importance of staff training was discussed in the Robert Wood Johnson Report and three different methods were used across three hospitals. In Central Georgia training in data collection was given to new staff members as part of induction training. Training for all staff that collect data on the importance of self-identification and the importance of and uses of the data is given in Boston. In Seattle, members of staff working in the registration areas are quality reviewed; managers are able to identify staff who record a large number of unknowns or blanks and implement training to address these problems [11].

## 6.7 Data quality and completeness

Completeness of ethnicity data is an ongoing problem. It was listed as one of the problems encountered during ethnicity data collection by the healthcare professionals' survey. Also the lack of 'use' of the data was a main complaint from focus group participants, but 'use' of data depends on good quality and complete data. Reports based on incomplete or bad quality data can provide misleading results. Incidence of renal disease in American Indians/Alaska Natives increased from 268 per million to 312 per million after corrections to coding; this has a profound effect on policies and provisions [2]. Ethnicity reported on data on death certificates was validated against data in the Indian Health Services database, 87% were found to be correct [3]. Ethnicity data validated across the 3 sources for AIDS patients, results show 57% conflicts in American Indians/Alaska Natives [30].

Many studies have compared self-reported data to official statistics and found inaccuracies [22]; it is important to have better data quality based on self-reported data. Ethnicity data was assessed in 376 recently diagnosed patients, findings showed that medical records are well linked to self defined ethnicity [31].

Incompleteness of ethnicity data is a major problem for UK cancer registration as they depend on third parties to provide this data. Jack et al (2006) reported that only 23% of registry data has ethnicity compared to 81% of HES data, and that linkage of records would be helpful to reduce duplication of work [32]. In the USA, a SEER initiative to assess the completeness of country of birth data reported that only 67% had recorded data with completeness varying by ethnic group, suggesting bias in collection [33]. As a consequence country of birth should be used with caution for surveillance and reporting purposes.

The Centres for Disease Control (CDC) observed no improvement in 'race' data collection between 1994 and 1997 [34]. In 1995 it became UK government policy to collect ethnicity data in a secondary care setting through HES. HES data collection has improved in London since the first collection in 1996/97 from 48% complete data to 65% complete data in 2001/02 [35, 36].

The importance of data collection is being recognised with new DoH initiatives for GPs to collect ethnicity for newly registered patients (1 QOF point). However there is a long way to go before databases hold complete and self-validated ethnicity data. The Lambeth PCT project demonstrates that with dedicated resources, training and monitoring, improvements can be made and awareness increased.

## 7 Conclusions

Incomplete data, conflicting data and lack of validation demonstrate the limited progress towards achieving a national policy of collecting ethnicity data. Some areas where initiatives have been assertively put in place, e.g. Lambeth PCT, have realised a significant improvement in data completeness and quality [10]. Other areas where there are a low population of ethnic minorities, and where it is not deemed to be important data, should still be encouraged to collect and report these data in order for policy makers to be able to determine high risk groups and inequalities on a national level.

There is a need to increase awareness about the importance of routinely collecting ethnicity. Ideally ethnicity should be collected as mandatory at the GP reception level as a self-reported field which is subsequently validated by discussion with the GP, with an opt-out 'not stated' option for those patients who refuse to provide their ethnicity once asked. It is well known that non-English speaking patients will often register with a same language speaking GP, thus making this an ideal setting for self-reported data collection and validation for those ethnic minorities with language barriers. Collection through the GP for all newly registered patients as well as self-reported ethnicity for existing patients may help to improve ethnicity data collection. It can be also collected at first hospital visit. However, ideally databases could be linked between primary and secondary care systems so that demographic data is collected once only, with validation thereafter. Olatokunbo et al (2000) showed successful collection of ethnicity data in a primary care feasibility study, they also demonstrated the ease with which ethnicity could be included on hospital referral letters by means of an automated field [37].

Many projects are ongoing – e.g. PROCEED - which aims to provide training to GPs and hospital staff about engaging with ethnic minorities and cultural awareness. Other training such as the NHS for Scotland toolkit [28] and Department of Health training developed in conjunction with the 2005 guidelines are resources which can be used to raise awareness and improve quality and completeness of ethnic data collection ([http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/StatisticalCollection/DH\\_4049499](http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/StatisticalCollection/DH_4049499)).

At the cancer registration level, identification of high risk groups can only be based on the current data collected. If this data is not available, poorly collected and remains not validated, then subsequent reports are unreliable.

## 8 References

1. Ma, I.W.Y., et al., *Systematic review identified suboptimal reporting and use of race/ethnicity in general medical journals*. *Journal of Clinical Epidemiology*, 2007. **60**(6): p. 572-578.
2. Sugarman, J.R. and L. Lawson, *The effect of racial misclassification on estimates of end-stage renal disease among American Indians and Alaska Natives in the Pacific Northwest, 1988 through 1990*. *American Journal of Kidney Diseases*, 1993. **21**(4): p. 383-6.
3. Frost, F., et al., *Correctness of racial coding of American Indians and Alaska Natives on the Washington State death certificate*. *American Journal of Preventive Medicine*, 1994. **10**(5): p. 290-4.
4. Swallen, K.C., et al., *Predictors of misclassification of Hispanic ethnicity in a population-based cancer registry*. *Annals of Epidemiology*, 1997. **7**(3): p. 200-6.
5. Gardi, M., *Ethnicity data protocols training presentation*. 2003, Ministry of Health Manatu Hauora.
6. *UK Census 2001*. [cited; Available from: <http://www.statistics.gov.uk/census2001/census2001.asp>].
7. *New federal standards for racial and ethnic data collection and reporting. A briefing on emerging issues in higher education from the Association for institutional research*. 1998 [cited; Available from: <http://www.fsu.edu/-air/alert6.htm>].
8. *Ethnicity: A Review of Data Collection and Dissemination*. 2003, Social and Housing Statistics Section, Demographic and Social Statistics Branch, United Nations Statistics Division.
9. *A Practical Guide to Ethnic Monitoring in the NHS and Social Care*. 2005, Department of Health and Social Care Information Centre/NHS Employers.
10. *Lambeth PCT peer review*. 2006, Race for Health.
11. Regenstein M and Sickler D, *Race, ethnicity, and language of patients. Hospital practices regarding collection of information to address disparities in health care*. 2006, The Robert Wood Johnson Foundation and National Public Health and Hospital Institute.
12. Hasnain-Wynia R, Pierce D, and Pittman MA, *Who, when, and how: The current state of race, ethnicity, and primary language data collection in hospitals*. 2004, Health Research and Educational Trust and The Commonwealth Fund.
13. Weinick, R.M., et al., *Measuring racial and ethnic health care disparities in Massachusetts*. *Health Affairs*, 2007. **26**(5): p. 1293-1302.
14. Ford, M.E. and P.A. Kelly, *Conceptualizing and categorizing race and ethnicity in health services research*. *Health Services Research*, 2005. **40**(5 Pt 2): p. 1658-75.
15. *Collecting ethnic category data: Guidance and training material for implementation of the new ethnic categories*. 2001, Department of Health.
16. *Ethnic monitoring: A guide for public authorities*. 2002, Commission for Racial Equality.

17. *Ethnic group statistics: A guide for the collection and classification of ethnicity data*. 2003, A National Statistics publication.
18. Adebayo C and Mitchell P, *Patient profiling*. 2004, Central Liverpool Primary Care Trust: Liverpool.
19. Gill PS, et al., *Black and minority ethnic groups*, in *Healthcare Needs Assessment*.
20. Hasnain-Wynia, R. and D.W. Baker, *Obtaining data on patient race, ethnicity, and primary language in health care organizations: Current challenges and proposed solutions*. Health Services Research, 2006. **41**(4): p. 1501-1518.
21. *HRET Disparities Toolkit: A toolkit for collecting race, ethnicity and primary language information for patients*. 2007 [cited].
22. Buescher, P.A., Z. Gizlice, and K.A. Jones-Vessey, *Discrepancies between published data on racial classification and self-reported race: evidence from the 2002 North Carolina live birth records*. Public Health Reports, 2005. **120**(4): p. 393-8.
23. Chattar-Cora, D., et al., *Colorectal cancer in a multi-ethnic urban group: its anatomical and age profile*. International Surgery, 2000. **85**(2): p. 137-42.
24. Sheth, T., et al., *Classifying ethnicity utilizing the Canadian Mortality Data Base*. Ethnicity & Health, 1997. **2**(4): p. 287-95.
25. Warnakulasuriya, K.A., et al., *Cancer of mouth, pharynx and nasopharynx in Asian and Chinese immigrants resident in Thames regions*. Oral Oncology, 1999. **35**(5): p. 471-5.
26. Baker, D.W., et al., *Patients' attitudes toward health care providers collecting information about their race and ethnicity*. Journal of General Internal Medicine, 2005. **20**(10): p. 895-900.
27. Baker, D.W., et al., *Attitudes toward health care providers, collecting information about patients' race, ethnicity, and language*. Medical Care, 2007. **45**(11): p. 1034-1042.
28. *Ethnic Monitoring Tool*. 2005 [cited; Available from: <http://www.isdscotland.org/isd/files/ETHNIC%20MONITORING%20TOOL.pdf>].
29. *The PROCEED project*. [cited; Available from: <http://info.cancerresearchuk.org/proceed/>].
30. Kelly, J.J., et al., *Race/ethnicity misclassification of persons reported with AIDS. The AIDS Mortality Project Group and The Supplement to HIV/AIDS Surveillance Project Group*. Ethnicity & Health, 1996. **1**(1): p. 87-94.
31. Gotay, C.C. and J. Holup, *Ethnic identities and lifestyles in a multi-ethnic cancer patient population*. Pacific Health Dialog, 2004. **11**(2): p. 191-8.
32. Jack, R.H., et al., *Ethnicity coding in a regional cancer registry and in Hospital Episode Statistics*. BMC Public Health, 2006. **6**: p. 8.
33. Lin, S.S., C.D. O'Malley, and S.W. Lui, *Factors associated with missing birthplace information in a population-based cancer registry*. Ethnicity & Disease, 2001. **11**(4): p. 598-605.
34. Centers for Disease Control and P., *Reporting race and ethnicity data-- National Electronic Telecommunications System for Surveillance, 1994-1997*. MMWR - Morbidity & Mortality Weekly Report, 1999. **48**(15): p. 305-12.



35. *HESonline: How good is HES ethnicity coding and where do the problems lie?* 2004 [cited; Available from: <http://www.hesonline.nhs.uk>.
36. *Diversity counts. Ethnic health intelligence in London: the story so far.* 2003, London Health Observatory.
37. Olatokunbo, S. and R. Bhopal, *Can we implement ethnic monitoring in primary health care and use the data? A feasibility study and staff attitudes in North East England.* *Public Health Medicine*, 2000. **2**(3): p. 106-108.



## Appendix 1: CanEth and Advisory board members

### CanEth group

Prof Janet Dunn	University of Warwick
Prof Ala Szczepura	University of Warwick
Prof Mark Johnson	De Montfort/Warwick University
Prof Sue Wilson	University of Birmingham
Dr Anil Gumber	University of Warwick
Gulnaz Begum	University of Warwick

### Advisory board attendees

Chris Carrigan	National Cancer Registry Coordinator
Catherine Thomson	West Midlands Cancer Intelligence Unit
Veronica Nanton	University of Warwick
Mike McCleod	Northamptonshire County Council
Ronny Flynn	Race Equality Foundation
Christine Gratus	Consumer Liaison Group
Jackie Brown	Lilly, Global Health Outcomes Scientist
Ruth Yates	Cancer Research UK
Neera Deepak	PRIAE: Policy Research Institute on Ageing & Ethnicity
Rose Thompson*	BME Cancer Information Specialist

\*Could not attend Advisory Board but requested an information pack and final report

## Appendix 2: Search terms for literature search

data collection - 15

271106 11.59 am

1. (multicultural or multi-cultural).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
2. (crosscultural or cross-cultural).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
3. (transcultural or trans-cultural).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
4. (multiethnic or multi-ethnic).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
5. (multiracial or multi-racial).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
6. (migrant\$ or immigrant\$).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
7. refugee\$.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
8. cultural diversity.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
9. (multilingual or multi-lingual).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
10. (romany or romanes or gypsy or gypsies).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
11. asylum seeker\$.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
12. (arab\$ or somali\$ or yemini\$ or Vietnamese or chinese\$ or caribbean or pakistani\$ or indian\$ or bangladeshi\$).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
13. (Islam\$ or Hindu\$ or Sikh\$ or buddhis\$ or muslim\$ or moslem\$).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
14. mixed race\$.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
15. (ethnocultural or sociocultural).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
16. diverse population\$.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
17. ((Black or ethnic or minorit\$) adj5 population\$).ab,ti.
18. (BME and ethnic\$).ab,ti.
19. BME.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]
20. or/1-19
21. Pro-forma\$.ab,ti.
22. coding.ab,ti.
23. (record\$ and keep\$).ab,ti.
24. (data adj3 collect\$).ab,ti.
25. (ethnic\$ and (record\$ or profil\$ or monitor\$)).ab,ti.
26. or/21-25
27. 20 and 26
28. limit 27 to english language
29. limit 28 to humans
30. exp Diabetes Mellitus/
31. diabet\$.ab,ti.
32. exp Hypertension/
33. hypertension.ab,ti.
34. Coronary heart disease.mp. or exp Coronary Disease/
35. heart disease\$.ab,ti.
36. (CHD and heart).ab,ti.
37. exp Cerebrovascular Accident/
38. stroke\$.ab,ti.
39. exp neoplasms/
40. (cancer\$ or tumor\$ or tumour\$ or malignan\$ or oncolog\$ or carcinoma\$ or neoplasm\$).ab,ti.
41. long term disease\$.ab,ti.
42. Chronic disease\$.ab,ti.
43. disease\$.ab,ti.
44. or/30-43
45. 29 and 44
46. limit 45 to yr="1990 - 1999"\*\*\*
47. remove duplicates from 46

**\*\*Dates changed to “2000-2007” for 2<sup>nd</sup> search**

## Appendix 3: Ethnicity, Health and Diversity questionnaire

Name of organisation:

Position (circle as appropriate): Clinician / Manager / Nurse/ Information Scientist / Other

Job title:

Ethnicity data collection (this includes:- ethnic group, language, religion, country of origin, country of birth, racial category)

1. Please rate how important you personally think the collection of ethnicity data is on a scale of 1 to 5:

Unimportant 1	2	3	4	Very important 5
------------------	---	---	---	---------------------

2. Do you attempt to collect any ethnicity data on patients?  No /  Yes

2a. If ethnicity data are not collected please give reasons below and go to question 5

2b. For which disease areas do you routinely collect ethnicity data (please tick all relevant boxes):

All disease areas      
 Cancer      
 Diabetes      
 Hypertension   
 Other      
 If other please state:

2c. For the routine data collection indicated in 2b above, please estimate the overall % for which you have recorded ethnicity  %

3a. If ethnicity data are collected please state the methods used:

Patient self-assessment   
 Assessment by healthcare professional by observation   
 Other    
 If other please give details

e.g. Indirect assessment using country of origin or name recognition software

3b. Please comment on any problems you have encountered when collecting ethnicity data:

**3c. Which indicators of ethnicity do you routinely collect (please circle all relevant responses)?**

Census ethnic group	<b>No / Yes / Not known</b>	Country of birth	<b>No / Yes / Not known</b>
Race	<b>No / Yes / Not known</b>	Country of origin	<b>No / Yes / Not known</b>
Religion	<b>No / Yes / Not known</b>	Language	<b>No / Yes / Not known</b>
Patient name (i.e. for use with name recognition software)	<b>No / Yes / Not known</b>		
Other	<b>No / Yes / Not known</b>		
If other please give details	<input type="text"/>		

**4. Are you using any name recognition software (e.g. Nam Pehchan or SANGRA)?**

No (go to question 5)

Yes, please state which

**4a. What is your experience (in terms of reliability) of using such software?** **High / Medium / Low**

**4b. Have you compared the results of this software with other data sources?** **No / Yes / Not known**

**4c. Have you developed a local dictionary to enhance its reliability?** **No / Yes / Not known**

**5. If not used in the past, would you be interested in using name recognition software?** **No / Yes**

**6. Does your organisation provide any training in ethnic monitoring?** **No / Yes / Not known**

**7. Would you be interested in attending an 'ethnic monitoring & its uses in cancer' workshop?** **No/Yes**

**8. Please rate the value of collecting ethnicity data to your organisation:**

Unimportant				Very important
1	2	3	4	5

**9. Any other comments:**

Would you be prepared to speak to us about this area? If yes, please provide your contact details below:

Name: \_\_\_\_\_ Email: \_\_\_\_\_ Tel: \_\_\_\_\_

**Thank you very much for your patience in completing this questionnaire.**  
**Would you like to receive a copy of the final report: Yes / No**

## Appendix 4: Topic guide

### CanEth Project Topic Guide: Warwick, Birmingham & De Montfort Universities

**Object: to understand if people using health services are happy to describe their ethnic origin – if they have been asked, and if they understand why.**

Note: We are especially interested in certain diseases, because people from minority ethnic groups have different risks, and may not get equal access to services. We believe that asking about 'ethnic origin' may help us to understand things better.

**Introduction:** We live in a multicultural society: some people use labels about ethnic groups a lot, and some places ask you what your 'ethnic group' is, or 'where you are from', when you go to use their services.

#### General opinions on the collection of ethnicity

- Do you think accurate recording is important?
- What do you think it can be used for?
- Any objections/worries about providing this data?

#### Experiences of providing ethnicity information

- General discussion THEN Focus down on healthcare situations
- Does anyone know people who have been asked this in relation to 'health research'?
- Does anyone know if the Cancer clinics ask these questions?

#### Categories used in practice (provide examples on sheets)

- Census
- Hospital admissions
- GP data
- Other
- What categories would you like – how would you prefer to describe yourself

#### What about language, Religion, Culture:

- Do people ask, do you offer this information, do you mind.
- Are there problems with 'stereotypes' (Explain)

**How should this information be collected (if it has to be: Note – the 'Race Relations Act' says that public services should so they can 'combat ethnic inequality')**

- Would you recommend the routine collection at hospital/GP/other?
- When would be the best time to collect this data (admission/follow-up after you've been to the hospital once)?
- How should people ask you – and what should they tell you?
- Has anyone in the group been asked to take part in 'research' at the hospital or their GP? (i.e. medical research) – Can you tell us about it?

#### Closing comments

- Does it make a difference in the case of a disease like **cancer** – or is it the same for any health matter?

Is there anything else you want to tell us about?

## Appendix 5: NCRI poster 2007

### Improving ethnic data collection for statistics of cancer incidence, management, mortality and survival in the UK

G Begum, A Gumber, MRD Johnson, A Szczepura, S Wilson, JA Dunn

University of Warwick, UK, De Montfort University, Leicester, UK, University of Birmingham, UK

#### Background

- Accurate 'ethnicity' data is essential to inform policy makers, funders and public health experts of incidence, prevalence and outcomes of specific conditions in population subgroups
- Some ethnic minority groups are associated with increased incidence of diabetes, hypertension, stroke and certain cancers
- 4.5 million people (8% of population) within the UK in 2001 were defined as being from an ethnic minority group; South Asians accounting for 50% (figure 1)
- Reports suggest that both breast and colorectal cancers were lower in the South Asian population; however this is rapidly increasing over time
- UK Government initiatives are in place to collect ethnicity but are limited to hospital admissions data. However, the data remains incomplete and has not improved over time (figure 2) and where collected the accuracy of data collection has not been validated?

Figure 1

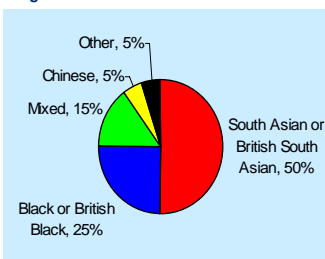
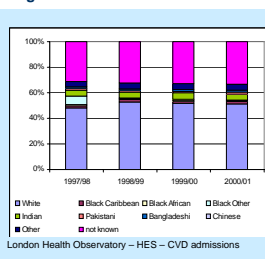


Figure 2



#### Aims

'CanEth' (Cancer Ethnicity project) is a feasibility project aiming to gather robust evidence, identify clear solutions and make recommendations to improve the collection of ethnicity data for UK cancer statistics

#### Objectives of 'CanEth'

- Literature review of methods, interventions and barriers addressing collection of ethnicity data in primary/secondary care
- Evaluation of health care professionals' perceptions and experiences of collecting ethnicity data in primary/secondary care
- Evaluation of consumers' perceptions, experiences and willingness to provide ethnicity data in primary/secondary care
- Validation (completeness and accuracy) of ethnicity data collected in a feasibility study of selected PCTs

#### Methods

- Literature searches were carried out using Embase, Psychlit, MEDLINE, PsycINFO, CINAHL and Google/Scholar
- Published literature was reviewed in 3 stages: 1) titles were coded as 'possibly relevant' or 'not relevant', 2) 'possibly relevant' titles advanced to the abstract stage where they were coded as 'relevant' or 'not relevant', 'relevant' abstracts were carried through to the manuscript stage 3) manuscripts were coded as either 'methodology' (data collection/monitoring) or 'use' (of ethnicity data).
- Questionnaires distributed through Minority-Ethnic-Health and ALLSTAT email groups and the National Cancer Research Network
- Focus groups formed from existing contacts with volunteer groups and facilitators to include main minority groups, i.e. African-Caribbean, Pakistani/Kashmiri (Muslim), Punjabi/Sikh (Indian), Gujarati (Hindu and Muslim), and Bangladeshi
- Applying the 'Nam Pehchan' name recognition software to 4 local general practice data bases will allow validation of South Asians

#### Results

##### Objective 1: Published literature search

Period	No of titles	No of abstracts	No of articles
1990-1999	722	5	5
2000-2007	1578	92	13
Total	2300	97	18

Category	1990-1999	2000-2007	Total
	No of articles	No of articles	No of articles
Methodology	5	6	11
Use	0	7	7
Total	5	13	18

##### Summary of published literature

- Paucity of published evidence regarding methodology of ethnicity data collection or ethnicity profiling
- Many articles use ethnicity data in their analyses of clinical data, health surveys or risk assessment of particular diseases
- Self completed ethnicity better than observer assessment
- Official ethnicity categories need to be re-examined and defined better

#### Grey literature

Searches carried out on Google ([www.google.co.uk](http://www.google.co.uk)) and Google Scholar (<http://scholar.google.com/>) using terms 'ethnicity', 'data collection' and 'monitoring' produced over a million results, many of which led to specialist websites

Searches carried out on specialist websites:

- Specialist Library for Ethnicity and Health (<http://www.library.nhs.uk/ethnicity/>)
- 'Race for Health' (<http://www.raceforhealth.org/>)
- Centre for Evidence in Ethnicity Health and Diversity
- Public Health Observatory/London Health Observatory (<http://www.lho.org.uk/>)
- Produced 12 key reports of which 5 are guidelines, 5 methods/use and 2 to be synthesised

#### Objective 2: Questionnaire on current practice

- 13 questionnaires returned to date (closing date 31<sup>st</sup> May 2007)
- Majority of people were from health care NHS trusts
- None had adequate training on the collection of ethnicity data
- None had validated the data

#### Reasons given why ethnicity data not collected

"It is very difficult to record ethnicity data for our cancer records as it is not documented in the patient's case notes, to the best of my knowledge. Due to this, it would take a great deal of time to collect and is however, not asked for in any reports that are asked of me" (Acting Cancer Services Coordinator)

"Ethnicity data collection currently limited to Midwifery as Trust is taking part in the Welsh Assembly Government Patient Equality Monitoring Project and staff are awaiting training in how to collect information" (Human Resources Manager)

"We have not to date regarded it as sufficiently important" (Consultant Gastroenterologist)

"Not relevant to care or treatment given to patients. York has very few ethnic groups therefore language diet etc not required. Would access if appropriate" (Sister Research Oncology)

"Only if it is required as part of a research trial and the company require that information. We then only fill it in, but it is very rare. We do not routinely collect this" (Lead Research Nurse)

#### Ongoing work

##### Objective 3: Focus groups

- Focus groups will be conducted with our local ethnic minority groups: African-Caribbean, Pakistani/Kashmiri (Muslim), Punjabi/Sikh (Indian), Gujarati (Hindu and Muslim), and Bangladeshi
- Focus group facilitators have experience in administration of qualitative 'topic guides' for group discussion; proposed topic guide for CanEth:

##### General opinions on the collection of ethnicity

- Do you think accurate recording is important?
- What do you think it can be used for?
- Any objections/worries about providing this data?

##### Experiences of providing ethnicity

- General discussion
- Focus down on healthcare situations

##### Categories used in practice (provide examples on sheets)

- Census
- Hospital admissions
- GP data
- Other

##### Closing comments

- Is there anything else you want to tell us about?
- Would you recommend the routine collection at hospital/GP/other?
- When would be the best time to collect this data (admission/fu)?
- Any final comments?

##### Objective 4: Data validation

Nam Pehchan and SANGRA (South Asian Names and Group Recognition Algorithm) name recognition software have been developed to identify South Asians. Previous work suggests that Nam Pehchan has a 91% sensitivity and 99% specificity<sup>2</sup>, whilst SANGRA has a 89-91% sensitivity and 94-98% specificity<sup>4</sup>.

Databases from 4 General Practitioners who serve a population of 'ethnic rich' patients across the West Midlands will be evaluated using the 'Nam Pehchan' software. This software will identify South Asians whose records will be cross-matched against their recorded 'ethnicity'. This will allow validation of the recorded ethnicity and provide some indication as to the accuracy of these data.

#### Conclusions

- There is a paucity of information on the best way to collect ethnicity
- When ethnicity is collected in UK healthcare systems, there is a lack of validation
- Focus groups will determine how minorities perceive the importance of data collection and compliance to the data recording
- Data Validation on local databases will determine accuracy of South Asians
- This is a feasibility study commissioned by Cancer Research UK

#### References

- Office for National Statistics <http://www.statistics.gov.uk/cci/nugget.asp?id=273>
- London health observatory <http://www.lho.org.uk>
- Cummins C, Winter H, Cheng KK et al. An assessment of the Nam Pehchan computer program for the identification of names of South Asian ethnic origin. Journal of Public Health, 1999;21: 401-6
- Nanchahal K, Manghani P, Alston M, dos Santos Silva I. Development and validation of a computerized South Asian Names and Group Recognition Algorithm (SANGRA) for use in British health-related studies. J Public Health Med. 2001 Dec;23(4):278-85

Email: [g.begum@warwick.ac.uk](mailto:g.begum@warwick.ac.uk)