

Improving ethnicity data collection for health statistics in the UK

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

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

<https://shura.shu.ac.uk/5567/>

This document is the Published Version [VoR]

Citation:

IQBAL, Gulnaz, GUMBER, Anil, JOHNSON, Mark R.D, SZCZEPURA, Ala, WILSON, Sue and DUNN, Janet A (2009). Improving ethnicity data collection for health statistics in the UK. Diversity and Equality in Health and Care, 6 (4), 267-285. [Article]

Copyright and re-use policy

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

Research paper

Improving ethnicity data collection for health statistics in the UK

Gulnaz Iqbal BSc (Hons)

Research Fellow, Warwick Clinical Trials Unit

Anil Gumber PhD

Senior Research Fellow, Centre for Evidence in Ethnicity, Health and Diversity

University of Warwick, Coventry, UK

Mark RD Johnson PhD

Professor of Diversity in Health and Social Care, Mary Seacole Research Centre, De Montfort University, Leicester, UK

Ala Szczepura DPhil

Professor of Health Services Research, Centre for Evidence in Ethnicity, Health and Diversity, University of Warwick, Coventry, UK

Sue Wilson PhD

Professor of Clinical Epidemiology, Department of Primary Care, University of Birmingham, Birmingham, UK

Janet A Dunn PhD

Professor of Clinical Trials and Head of Cancer Trials, Warwick Clinical Trials Unit, University of Warwick, Coventry, UK

What is known on this subject

- Disparities in health between ethnic groups have been widely reported.
- Ethnic record keeping/monitoring has been undertaken for the past 20 years, but has often been carried out in an ad-hoc manner, resulting in incomplete and unvalidated data.
- There is an urgent need to routinely collect and use good-quality ethnicity data in order to reduce health inequalities and target services appropriately.

What this paper adds

- Many clinical papers report the use of ethnicity data in their analyses, health surveys or risk assessments of particular diseases, but details of the methodology of data collection on ethnicity are often lacking or incomplete.
- Better understanding of the dimensions of ethnicity, using more satisfactory categories, will enhance understanding of the effects of ethnic group membership on outcomes.

ABSTRACT

There is an identified need for the collection of ethnicity data in the healthcare setting. Accurate data on ethnicity are essential for informing policy makers, funders and public health experts about the incidence, prevalence and outcomes of specific conditions in population subgroups. There is emerging evidence that some ethnic groups are associated with an increased incidence of certain cancers, and

disparities in access to services have been documented. Government initiatives are in place to collect ethnicity data in the healthcare setting, but the accuracy of the data needs to be validated.

Cancer Research UK commissioned the Cancer Ethnicity (CanEth) project to gather robust evidence and identify solutions to improve the collection of ethnicity data for cancer. The project set out to

review current literature focusing on methods, interventions and barriers addressing the collection of ethnicity data.

The review identified a paucity of published evidence on ethnicity data collection. Many clinical articles used ethnicity data, but few discussed the methodology of data collection. In general, however, self-reported ethnicity is recognised as the best

method of data collection, and is preferable to observer assessment. Training is needed to raise awareness of the importance of ethnicity data and its use to facilitate the reduction of inequalities.

Keywords: black and minority ethnic groups, data collection, ethnicity, monitoring, profiling

Introduction

The reduction of cancer inequalities was a key feature of the Cancer Reform Strategy published in 2007, which proposed to improve cancer outcomes and uptake of services by 2012, including those inequalities observed in black and minority ethnic (BME) populations (Department of Health, 2007). In cancer, ethnicity data collection and monitoring are particularly important because ethnic minority groups have been demonstrated to have later presentation, leading to poor survival (Smith *et al*, 1999; White, 2002). Also, some ethnic minority groups tend to demonstrate more risky behaviour. For example, smoking rates were reported to be highest in Bangladeshi males (44%), followed by Irish males (39%), compared with 27% in the general population, whereas Bangladeshi women are more likely to chew tobacco (26%) than to smoke cigarettes (White, 2002). Reports suggested that the incidence of both breast and colorectal cancer was lower in the South Asian population. However, incidence rates are increasing over time (Smith *et al*, 2003; Farooq and Coleman, 2005). With regard to other disease areas, South Asians in the UK are 50% more likely to die prematurely from coronary heart disease than the general UK population, and males and females of Pakistani and Bangladeshi origin are six times more likely than the general population to have diabetes (Townsend *et al*, 1988; Commission for Racial Equality, 2008).

The 2001 census classified 4.6 million people (7.9%) in the UK as belonging to a non-white ethnic group, with over 50% of these classified as Asian or British Asian (Office for National Statistics, 2001). This is an increase compared with the 5.5% of the population not defined as white in the 1991 census. The 2001 census identified 55% of the mixed race category as being 16 years of age or younger. For epigenetic modelling, a more detailed definition of 'mixed race' is required, such as mothers', fathers' and grandparents' ethnicity and geographical origins/ancestry. To improve public services appropriate to the needs of BME patients, 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 (Johnson, 1998, 2001, 2006; White, 2002; London Health Observatory, 2003; Greater London Authority, 2005). The quality of ethnicity data recording has been variable. Attempts to improve the completeness and quality require dedication and commitment (Liverpool John Moores University, 2000). Reports focusing 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 are limited (White, 2002; Greater London Authority, 2005). 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 regard to completeness and accuracy (Pringle and Rothera, 1996; Kumarapeli *et al*, 2006; Jones and Kai, 2007). 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) of individuals to provide information, and a lack of understanding of 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 record ethnicity accurately. The danger is that current policies are based on inaccurate data and, as such, may lead to inappropriate distribution of resources and services (White, 2002; London Health Observatory, 2003; Greater London Authority, 2005).

In 1995 it became UK government policy to collect ethnicity data in secondary care settings through Hospital Episode Statistics (HES). HES data collection has improved over time. For example, in London, 52% of records in 1996/1997 had incomplete data, whereas by 2001/2002 this figure had fallen to 35% (London Health Observatory, 2003).

In 2001–2002, an attempt was made to increase ethnicity profiling in primary care. However, at this time the work involved and the related costs were significant deterring factors (Jones and Kai, 2007). Recently, some primary care trusts have invested in the collection of ethnicity data, 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 one point) (Race for Health, 2007). Monitoring goals set for London for 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 expected this figure to reach 95% by March 2006 (London Health Observatory, 2003). The ‘Professionals Responding to Cancer in Ethnic Diversity’ (PROCEED) project team provided training in competence and cultural awareness for healthcare professionals who were involved in cancer care at primary care level. The issues explored included cancer and ethnic diversity, language and communication, and culture and cancer (Cancer Research UK, 2006).

In 2005, the NHS produced a guide to ethnic monitoring in the NHS and social care, with several examples of good practice (Department of Health, 2005). There is limited information on the uptake of these guidelines and their practical applicability. Within the cancer setting, family history, ethnicity, social class, material deprivation, lack of access to services and subsequent delay times have all been adversely linked to outcome (i.e. survival) (Townsend *et al*, 1988; White, 2002; Farooq and Coleman, 2005; Woods *et al*, 2006). 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.

This paper is the first part of a project commissioned by Cancer Research UK to assess ethnicity data collection for statistics of cancer incidence, management, mortality and survival in the UK. The report also includes a survey of healthcare professionals’ perceptions of ethnicity data collection, focus groups of consumers’ perceptions and willingness to provide ethnicity data in healthcare, and a validation exercise to assess the completeness and accuracy of ethnicity data in a feasibility study of GP practices (Iqbal *et al*, 2008).

This paper focuses on one part of the project, namely a systematic review undertaken to gather robust evidence and identify clear solutions and recommendations to improve the collection of ethnicity data for health statistics in the UK. This information is essential in order to obtain a better understanding of the uptake of services and health outcomes, to monitor

trends, to target interventions and allocate resources to better meet the needs of BME groups, and to tackle health inequalities. The review examined the published literature discussing methods, interventions and barriers with regard to the collection of ethnicity data in primary and secondary care. It also included a separate search of key websites to identify relevant ‘grey literature’ such as government reports and other unpublished material which cannot easily be found via conventional database searches.

Methods

Published literature

The databases used for this review were identified in the early stages of the project through consultation with a team of experts, including a specialist information scientist working for the Centre for Evidence in Ethnicity, Health and Diversity (CEEHD). The searches encompassed five bibliographic databases, namely Embase, Psychlit, Medline, Psycinfo and Cinahl. 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 and coronary heart disease (see Table 1). The search of published literature was split into two sections. The first search was limited to 2000–2007 with the aim of identifying recent literature. The second search used the same terms but was extended to 1990–1999 to capture literature before and after the National Institute of Health Revitalization Act, which was passed in the USA in 1993 and prompted interest in reporting by ethnic group. The review was conducted in three stages, namely title, abstract and article review. Abstracts were reviewed by the researcher and by the co-authors as well as by members of an independent advisory board of experts.

Grey literature

Grey literature searches were conducted using the keywords *data collection* OR *data monitoring* AND *ethnic* OR *ethnicity*. The searches were performed in Google and Google Scholar. Only the first 50 pages were scanned, due to the huge volume of results. In addition, extensive searches were carried out on key websites such as the Specialist Library for Ethnicity and Health, the London Health Observatory, the Office for National Statistics and the Department of Health.

The findings are presented in sections based on seven themes which emerged during the course of the review as shown in Box 1.

Table 1 Free text and MeSH indexing terms

Ethnicity	Disease sites	Data collection
1. (multicultural or multi-cultural).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	1. exp Diabetes mellitus/	1. Pro-forma\$.ab,ti.
2. (crosscultural or cross-cultural).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	2. diabet\$.ab,ti.	2. coding.ab,ti.
3. (transcultural or trans-cultural).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	3. exp Hypertension/	3. (record\$ and keep\$).ab,ti.
4. (multiethnic or multi-ethnic).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	4. hypertension.ab,ti.	4. (data adj3 collect\$).ab,ti.
5. (multiracial or multi-racial).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	5. Coronary heart disease.mp. or exp Coronary disease/	5. (ethnic\$ and (record\$ or profil\$ or monitor\$)).ab,ti.
6. (migrant\$ or immigrant\$).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	6. heart disease\$.ab,ti.	
7. refugee\$.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	7. (CHD and heart).ab,ti.	
8. cultural diversity.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	8. exp Cerebrovascular accident/	
9. (multilingual or multi-lingual).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	9. stroke\$.ab,ti.	
10. (romany or romanies or gypsy or gypsies).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	10. exp neoplasms/	
11. asylum seeker\$.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	11. (cancer\$ or tumor\$ or tumour\$ or malignan\$ or oncolog\$ or carcinoma\$ or neoplasm\$).ab,ti.	
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]	12. long term disease\$.ab,ti.	
13. (Islam\$ or Hindu\$ or Sikh\$ or buddhis\$ or muslim\$ or moslem\$).mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	13. Chronic disease\$.ab	
14. mixed race\$.mp. [mp=ti, hw, ab, it, sh, tn, ot, dm, mf, nm]	14. disease\$.ab	
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]		

Box 1 Topics to be addressed by the studies

- Ethnicity data collection and monitoring
- Categories for defining ethnic group
- Other indicators of ethnicity
- Methods of data collection
- Barriers to data collection
- Interventions
- Data quality and completeness

Results

Overview

The majority of the relevant published articles were from the USA (68%). However, the majority of guidelines found in the grey literature search were UK based (63%). Of the 35 articles included in the review, 19 articles (54%) were identified from published literature and a further 16 articles (46%) from grey literature. In total, 29 (83%) of the relevant documents were interested in all ethnic groups, with six (17%) focusing on particular groups; 26% of the relevant literature consisted of either guidelines, training materials or toolkits.

Published literature

The review of published literature provided a total of 2404 'hits', of which 720 were for the period 1990–1999 and 1684 were for the period 2000–2007. 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 (see Figure 1). The main reason for rejection (57% of cases) was that the paper was concerned with the use of ethnicity data rather than the methods for collection of such data. The full text of the 26 potential articles was reviewed, and only 19 of these articles included information about data collection or monitoring. One of the potentially relevant papers is included based on the abstract only, as the full paper is unavailable (Chattar-Cora *et al*, 2000) (see Figure 1 and Table 2).

Grey literature

Searches on key websites and Google and Google Scholar identified a wealth of information, with 53 reports being identified as possibly associated with ethnic data collection or monitoring. The main reasons for rejection were that the reports contained only opinion (i.e. discussion of the need for ethnicity data collection) or used ethnicity data for reporting outcomes. Of the 53 reports that were reviewed, 16 were included in this review (see Table 3).

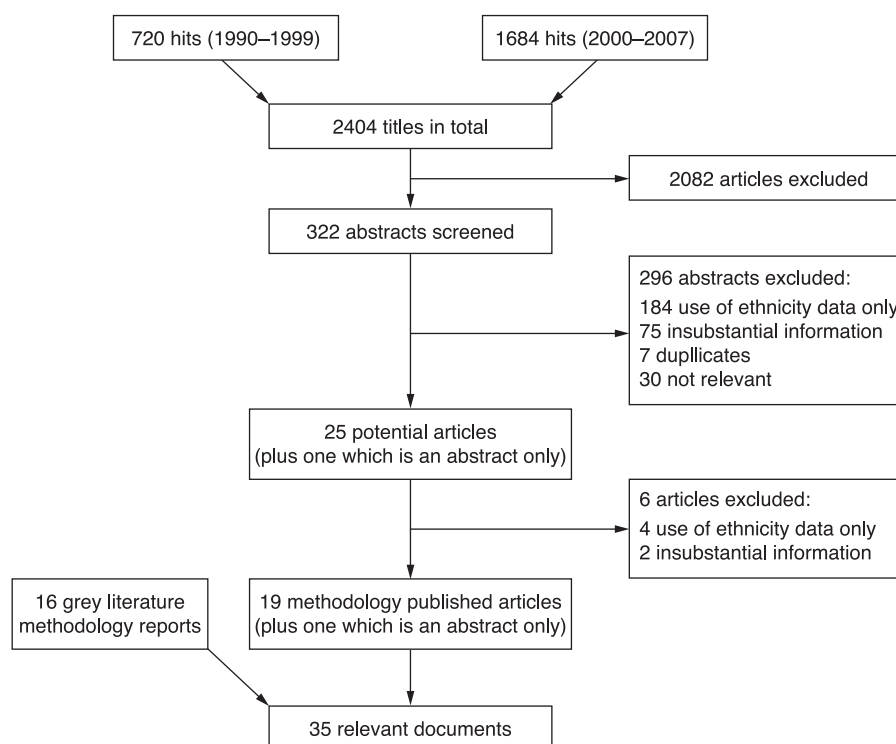


Figure 1 Ethnicity data collection and monitoring review selection process.

Table 2 Summary of published articles

First author, year of publication	Type of cancer	Country of study	Ethnic group	Type of study	Description of content	
					Focus of the study	Key findings
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 that the data should be collected; 46% were worried that the information would be used to discriminate against them; 17% were not comfortable reporting their own ethnicity
Ma, 2007	Non-cancer- specific	All	All	Systematic review	Methods of reporting race in medical journal articles	116 terms were used to describe ethnic groups; only 13% reported data collection method (1152 articles)
Weinick, 2007	Non-cancer- specific	USA	All	Review	New enactment of ethnicity data collection in acute care hospitals. Lessons learned from implementing publicly mandated data collection	Implementation of a change of policy needs to map on to existing systems, be flexible and be standardised. Training-for-trainers central sessions proved successful. Patient engagement and emphasis on the importance of data collection for improvements of care
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
Jack, 2006	All	UK	All	Audit	To determine the completeness of ethnicity data in Thames Cancer Registry and HES data held by London Health Observatory	81% of HES data had ethnicity recorded, compared with 23% in the registry. Better collaboration is needed between sources in order to improve registry ethnicity data

Table 2 Continued

Baker, 2005	Non-cancer- specific	USA	All	Cross-sectional	Patients' attitudes towards healthcare providers collecting ethnicity data	Patients are more willing to provide ethnicity data when the reasons for data collection are explained by staff in an appropriate manner. Staff should be comfortable collecting these data
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
Ford, 2005	Veteran Affairs	USA	All	Review	The importance of conceptualising and categorising ethnicity data	Better and more consistent methods of ethnicity data collection need to be developed
Gotay, 2004	All	Hawaii	Japanese Hawaiian European Filipino	Cross- sectional	To assess ethnic self-identity in 367 recently diagnosed ethnic patients, and to explore acculturation	Findings show that medical records are well linked to individual self-reported ethnicity
Lin, 2001	All	USA	All	Audit	SEER initiative to assess the completeness of data on country of birth	67% of patients on the register had birthplace recorded. Completeness of data varied between ethnic groups, suggesting that there was bias in collection of this item
Chattar-Cora, 2000 (abstract only)	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
Olatokunbo, 2000	Non-cancer- specific	UK	All	Feasibility study	Feasibility study of ethnic monitoring in primary care	Ethnic monitoring is feasible in primary care. The inclusion of ethnicity as an automated field on GP referral letters was shown to be a simple yet powerful method which can be used to populate hospitals' databases

Table 2 Continued

First author, year of publication	Type of cancer	Country of study	Ethnic group	Type of study	Description of content	
					Focus of the study	Key findings
Centers for Disease Control, 1999	Non-cancer	USA	All	Report	To assess the collection of race data in health surveillance systems between 1994 and 1997	No improvement in race data collection was observed between 1994 and 1997
Warnaku-lasuriya, 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
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 is more accurate than using country of birth alone
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 that Hispanics over-reported for 38% of cases. It recommends using both recorded ethnicity and name for increased accuracy
Kelly, 1996	Non-cancer, AIDS	USA	All	Audit	Validation of ethnicity classification for AIDS patients across three national data sources	Inconsistencies were greatest for Native Americans and Alaska Natives (up to 57% disagreement)
Frost, 1994	Non-cancer	USA	Native American Alaska Native	Audit	To validate race on Washington State death certificates with those on the Indian Health Service database	Race was correct for 87% of death certificates. Deaths from cancer were more likely to be coded incorrectly. People who are born and die in Washington are more likely to be coded correctly
Sugarman, 1993	Non-cancer, End-stage renal disease	USA	Native American Alaska Native	Audit	Misclassification of Native Americans and Alaska Natives on the Renal Disease Stage Register, and the impact upon disease statistics	Ethnicity was validated against the Indian Health Service database using names, date of birth and social security numbers. The incidence of renal disease increased from 268 per million to 312 per million after corrections to ethnicity coding

Table 3 Summary of grey literature reports

Title, year of publication	Authors	Country of report	Ethnic groups	Type of research	Description of content	
					Focus(es) of report	Key findings
HRET Disparities Toolkit: a toolkit for collecting race, ethnicity and primary language information for patients (amended version), 2007	Health Research and Educational Trust (HRET)	USA	All	Online toolkit	Designed to help healthcare workers to understand the importance of collecting good-quality data on ethnicity, race and preferred language	Toolkit includes the following 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 the race, ethnicity and primary language data to improve quality of care 7. How to train staff to collect this information
Lambeth Primary Care Trust review, 2006	Race for Health	UK	All	Paper	How successful Lambeth Primary Care Trust is in collecting, recording, analysing and using ethnicity monitoring information	Good practice includes the followig: 1. Individual Patient Registration Profile (IPRP), started in 2002, now with over 30 practices taking part. IPRP includes collection of data on religion, language, need for interpreter and ethnicity, as well usual data. Existing patients are contacted by postal questionnaire 2. Training for practice staff 3. DataNet system aids the use of collected data

Table 3 Continued

Title, year of publication	Authors	Country of report	Ethnic groups	Type of research	Description of content	
					Focus(es) of report	Key findings
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 to data collection	Overall collection of data is good, but the data are not put to use. Some confusion between ethnicity and race. Single most important barrier to data collection is staff not knowing why the data are important. Examples of good practice include: 1. training given to new staff members as part of induction 2. training for all staff collecting data on the importance of self-identification and uses of data 3. members of staff working in registration areas are subjected to quality review 4. managers are able to identify staff who record a large number of unknowns or blanks
Black and minority ethnic groups	Gill, Kai, Bhopal <i>et al</i>	UK	All	Needs assessment	A needs assessment overview for black minority ethnic groups (BMEGs) in the UK. Part of needs assessment series	No differences were reported in the rate of minority groups consulting their GPs or being 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-quality ethnic data in secondary care services

Table 3 Continued

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
Ethnic Monitoring Tool	NHS Scotland/ Health Scotland	UK	All	Toolkit	The tool has been designed for NHS Scotland staff involved in the collection or use of ethnicity data	Explains the need for data monitoring, who should be involved, and what needs to be in place, and provides some training materials
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	USA	All	Report	Survey and site visits to hospitals nationwide, to report current practice and identify problems	Reports inconsistencies in methods of data collection, questions asked and response categories. The report makes five recommendations: 1. Standardise the method of collection (self-report should be used whenever possible) 2. Point of data collection (admission) 3. Standardise the categories (ideally US Census) 4. Data storage should be standardised (e.g. race and ethnicity stored as two separate variables) 5. Response to patient concerns and explanations should also be standardised
Ethnicity data protocols training presentation, 2003	Gardi M. Ministry of Health, Manatu Hauora	New Zealand	All	Training presentation	Ethnicity data protocols; how to collect, classify and use ethnicity data	Ensure that ethnic groups of policy importance are not swamped by NZ ethnic group. Each patient only appears once, so the sum of the population adds up to NZ population. Advises against transferring ethnicity from one form to another

Table 3 Continued

Title, year of publication	Authors	Country of report	Ethnic groups	Type of research	Description of content	
					Focus(es) of report	Key findings
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 that 107 questions were asked by 95 countries. These can be placed in five categories: 43% of questions used a form of tick-box categories with an open-ended box for 'other', 20% had tick-box categories only, 21% were open-ended questions, 4% had yes or no responses, and 12% did not provide enough information
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	Two methods are proposed, namely one-question (ethnicity) and two-question (ethnicity and nationality) method. Two- question method should be used whenever possible
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 was identified as the poorest area, routine systems/integrated patient record could be possible solutions
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. Suggests that the method of collection should also be recorded

Table 3 Continued

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
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 is a legal requirement for all federal agencies. Self-identification should be used whenever possible. Proposes a two-question method for self-reports and a single-question method for data collection by observation
Patient profiling in primary care: the Princes Park Health Centre model, 2000	Liverpool John Moores University	UK	All with specific reference to Somali and Yemeni communities	Report	Reporting of patient profiling in primary care following the implementation of a Service Development initiative	Patient profiling data collected through the development and use of a Patient Information Form broken down into four sections (personal details, patient satisfaction, health and ill health, and ethnic classification). The data have been used to inform planning strategies, detailed in the report
Hospital Episode Statistics online, 2004	NHS	UK	All	Report	Ethnicity coding in HES:1997–98 to 2002–03	Overall records with missing ethnic data have decreased in the most recent 5-year period

Ethnicity data collection and monitoring

Six reports presented best practice evidence for ethnicity data collection and monitoring (Commission for Racial Equality, 2002; Department of Health, 2005; Health Scotland, 2005; Race for Health, 2006; Regenstein and Sickler, 2006; Health Research and Education Trust, 2007). Examples of best practice in the UK are given in the report by the Department of Health (2005). Key reports where ethnicity data collection has been successful due to adequate resources, awareness and training (Race for Health, 2006; Regenstein and Sickler, 2006) also demonstrated the need to have a 'use' for the data in order to improve collection.

Recommendations for improving ethnicity data collection are largely concerned with standardisation of the method of collection, point of collection, ethnicity categories, data coding and storage, and lastly standardised responses to the patients' frequently asked questions (Hasnain-Wynia *et al*, 2004; Ford and Kelly, 2005; Weinick *et al*, 2007). The UK Department of Health 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 (Department of Health, 2001; Hasnain-Wynia *et al*, 2004).

Categories for defining ethnic group

A United Nations report identified a total of 107 ethnicity questions asked by 95 countries in the census (United Nations Statistics Division, 2003). Only 12% of countries that collected ethnicity data had categories for 'mixed identities' or allowed multiple box selection. Other international guidelines indicate that the gold standard categories used within a country may be expanded so long as they can be concatenated back for national reporting purposes (Commission for Racial Equality, 2002; Race for Health, 2006; Weinick *et al*, 2007). There are also inconsistencies with the data types being used. These include coded tick box categories with and without boxes for free text, closed questions with yes/no responses and open questions for free text allowing people to describe themselves in their own words (United Nations Statistics Division, 2003).

The UK gold standard ethnicity categories are taken from the 2001 census ethnicity question which consists of 16+1 categories ('+1' being the code for 'not stated'). The Commission for Racial Equality (CRE) report and the Department of Health guide to ethnic monitoring both state the importance of not offering patients this option (Commission for Racial Equality, 2002; Department of Health, 2005).

Other indicators of ethnicity

The UK Department of Health guidelines encourage the additional collection of data on religion, diet, language and the need for an interpreter (Department of Health, 2005). These additional indicators of ethnicity should be collected especially if they are relevant at a local level. The Office for National Statistics (ONS) recommends that data on nationality are also collected for planning and resource purposes (Office for National Statistics, 2003). Responses should be re-ordered depending on where the question is being asked (e.g. in England, 'English' should be at the top of the list). This ordering to emphasise groups of policy importance is also practised in other countries, such as New Zealand, where 'Maoris' is at the top of the coding list (Gardi, 2003).

The Individual Patient Registration Profile (IPRP) used by Lambeth Primary Care Trust collects data on 'religion', 'language' and 'need for an interpreter' in addition to 'self-reported ethnicity' (Race for Health, 2006). The ethnicity categories have been expanded in line with the make-up of the local population, but can be concatenated to the census categories. The data are stored on a dedicated central database which can link the IPRP data to research projects. Central Liverpool NHS Primary Care Trust has also carried out patient profiling by collecting detailed ethnicity data, including 'spoken language' and 'reading language' (Liverpool John Moores University, 2000). However, 'country of birth', which has been collected since 1841, is no longer deemed a reliable indicator of ethnic origin, as at least 50% of members of ethnic minorities are born in the UK (Gill *et al*, 2007).

Methods of collection

Self-reported ethnicity is the gold standard, and the reasons for this are discussed in many good practice guidelines and papers (Commission for Racial Equality, 2002; Department of Health, 2005; Regenstein and Sickler, 2006). If healthcare professionals determine ethnicity by observation, this can lead to stereotyping by skin colour and name, so it should only be used where self-reporting is not possible. In the USA the Health Research and Educational Trust toolkit and Hasnain-Wynia *et al*. (2004) illustrate how staff should ask for these data, and emphasise the need for self-reporting (Hasnain-Wynia and Baker, 2006; Health Research and Educational Trust, 2007). Surveys conducted by the Robert Wood Johnson group showed that 61% of respondents usually asked the patient to self-report, but 25% filled in the ethnicity themselves on the basis of observation (Regenstein and Sickler, 2006). They felt that this method was easier for both them and the patient as it avoided any discomfort. They also felt that it was accurate, as they believed they

knew their local population. It would be informative to separate the occasions when staff fail to ask from those when patients do not wish to provide the data; these areas will need to be tackled independently, as they stem from different problems (Department of Health, 2005). The method of collection should also be recorded alongside the data (i.e. self-reporting or observation), otherwise other important biases could occur if assumptions are made about the reporting method (Commission for Racial Equality, 2002; Buescher *et al*, 2005). Sugarman and Lawson (1993) demonstrated that racial disparity varied according to the 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 the coding.

Other methods of collection could include the use of name recognition software. Patients' notes were used to successfully identify most patients in one study, demonstrating that names can be used with some precision when no other data are available (Chattar-Cora *et al*, 2000). It has been shown that name recognition software used in conjunction with other indicators such as country of birth results in increased accuracy (Sheth *et al*, 1997; Swallen *et al*, 1997; Warnakulasuriya *et al*, 1999).

Barriers to data collection

The main barrier to ethnicity data collection is staff members' lack of knowledge about the importance and use of the data. Site visits to six consortium member hospitals in the USA and a nationwide survey of 1000 hospitals found that 30% of respondents reported problems with or barriers to collecting ethnicity data (Hasnain-Wynia *et al*, 2004). The barriers reported were similar to those found in the Robert Wood Johnson report (Regenstein and Sickler, 2006), the most important being the reluctance of staff to ask for ethnicity data, due to fear of offending the patient or encountering resistance. Confusion about ethnicity categories, lack of a demonstrated need to collect the data, limitations of databases with regard to capturing this type of data, lack of resources, and lack of agreement among executive leaders about the need to collect these data were also reported (Hasnain-Wynia *et al*, 2004).

One of the main barriers to data collection is patients' perceptions. Baker reported that 46% of patients were concerned that the data would be used to discriminate against them (Baker *et al*, 2007). Patients would be more willing to provide data if the reasons why the data were being collected were explained to them, and healthcare professionals should be comfortable asking for these data (Baker *et al*, 2005).

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 (Commission for Racial Equality, 2002; Department of Health, 2005; Health Scotland, 2005; Race for Health, 2006; Regenstein and Sickler, 2006; Health Research and Educational Trust, 2007). The 2005 NHS guidelines state that staff training should be tailored to local need and should explain why ethnic monitoring is important, how to collect the data and what they will be used for. Local community groups could be asked to comment on the content of the training packs. All staff who may be involved in collecting ethnicity data, writing reports, or analysing or making decisions based on the data need to attend training. Training needs may differ from one group to another (Department of Health, 2005).

In the USA, the Health Research and Educational Trust toolkit provides a free national training package for the collection of ethnicity data (Health Research and Education Trust, 2007). It is written for all levels of healthcare workers, including chief executive officers, clinicians, registration staff and database managers, as well as for patients, enabling users to select the information package that is most relevant to them. The toolkit explains the need for ethnicity data collection, the need for standardisation, how to ask the questions, training exercises and how the data are or could be used. The resources provided include training presentations, definitions of key terms, and a reference booklet for staff.

Apart from the best practice guidelines in the UK, the most comprehensive training package is the Ethnic Monitoring Tool developed by NHS Scotland (Health Scotland, 2005). This is aimed at NHS Scotland staff and provides information on why it is important to carry out ethnic monitoring, who is involved, and what needs to be put in place. Training materials can be downloaded and modified according to local needs. Training-for-trainers notes and role-play scenarios are also provided. The Lambeth Primary Care Trust project offers 1.5 days of training for staff, computer templates are provided, and resources are made available to mail a questionnaire to existing patients as well as collecting ethnicity data for those newly registered (Race for Health, 2006).

The importance of staff training was discussed in the Robert Wood Johnson Report, with different methods used across three hospitals. The training was delivered as part of the induction programme to all new staff in the first hospital, but was provided to all staff in the second hospital. The third hospital subjected members of staff working in the registration areas to quality review. Managers are able to identify staff who record

a large number of unknowns or blanks, and implement training to address these problems (Regenstein and Sickler, 2006).

Quality and completeness of data

Completeness of ethnicity data is an ongoing problem. Reports based on incomplete or poor-quality data can provide misleading results. Many studies have compared self-reported data with official statistics and found inaccuracies (see, for example, Frost *et al*, 1994; Kelly *et al*, 1996; Buescher *et al*, 2005). It is important to have better data quality based on self-reported data. Ethnicity data were assessed in 376 recently diagnosed patients, and the findings showed that medical records are closely linked to self-defined ethnicity (Gotay and Holup, 2004).

Incompleteness of ethnicity data is a major problem for UK cancer registration, as registries depend on third parties to provide these data. Jack *et al* (2006) reported that ethnicity was recorded for only 23% of registry data, compared with 81% of HES data, and that linkage of records would be helpful to reduce duplication of work. In the USA, a Surveillance, Epidemiology and End Results (SEER) programme initiative to assess the completeness of data on country of birth reported that only 67% had recorded data, with completeness varying according to ethnic group, which suggests that there was bias in collection (Lin *et al*, 2001). Therefore country of birth should be used with caution for surveillance and reporting purposes.

The Centers for Disease Control (CDC) observed no improvement in race data collection between 1994 and 1997 (Centers for Disease Control and Prevention, 1999). However, an improvement has been seen in UK ethnicity data collection in secondary care since its inception in 1995 (London Health Observatory, 2003; Hospital Episode Statistics online, 2004). The importance of data collection is being recognised, but there is a long way to go before databases hold complete and self-validated ethnicity data. The Lambeth Primary Care Trust project demonstrates that, with dedicated resources, training and monitoring, improvements can be made and awareness increased.

Discussion and conclusion

This review has shown a need to increase awareness about the importance of routinely collecting ethnicity data. 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 when

asked to do so. 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 members of ethnic minorities with language barriers. Data 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. Ethnicity data can also be collected at the first hospital visit. However, ideally databases could be linked between primary and secondary care systems so that demographic data are collected once only, with validation thereafter. Olatokunbo and Bhopal (2000) showed successful collection of ethnicity data in a primary care feasibility study, and also demonstrated the ease with which ethnicity could be included on hospital referral letters by means of an automated field. Linkage of ethnicity data from the UK census with health databases has also been demonstrated to be tangible in a retrospective cohort study that explored variations in myocardial infarction in South Asians (Fischbacher *et al*, 2007).

Ethnicity has been an optional data item in Cancer Registry datasets since 1993, and has been poorly recorded, with many patients coded as 'not known.' Incomplete data, conflicting data and lack of validation demonstrate the limited progress towards achieving a national policy for collecting ethnicity data. At the cancer registration level, identification of high-risk groups can only be based on the current data collected. If these data are not available, poorly collected or remain unvalidated, subsequent reports will be unreliable. It is also important for collected data to be used when reporting outcome measures such as access to healthcare and uptake of services, and to feed into policies designed to tackle inequalities (Raleigh, 2008). Use of these data in such reports is needed to demonstrate the importance of collection to both patients and healthcare professionals.

Aspinall (2009) predicts increased complexity as categories for collecting ethnicity data are expanded in order to better capture the increasingly diverse population of the UK. This will include the addition of new items, such as 'national identity', which aim to further capture the multi-dimensionality of ethnicity. These changes will lead to increasing difficulties in the analysis of these data, but will allow the identification of groups with more than one identity (e.g. British Muslims), which has not been possible in the past (Aspinall, 2009).

Projects such as PROCEED (Cancer Research UK, 2006) aim to provide training for GPs and hospital staff about engaging with ethnic minorities and cultural awareness. Other training, such as the NHS Scotland toolkit (Health Scotland, 2005) and the Department of Health training that was developed in conjunction with the 2005 guidelines, offers resources which can be

used to raise awareness and improve the quality and completeness of ethnic data collection.

Some areas where initiatives have been assertively put in place (e.g. Lambeth Primary Care Trust, the Princes Park Health Centre and selected NHS boards in Scotland) have realised a significant improvement in data completeness and quality (Liverpool John Moores University, 2000; Race for Health, 2006; Information Services Division Scotland, 2009). Other areas where there is a low population of ethnic minorities, and where ethnic diversity is not deemed to be locally significant, should still be actively encouraged to collect and report these data in order to enable policy makers to determine high-risk groups and inequalities at a national level. It is imperative that the current levels of national awareness and motivation with regard to the importance of ethnic data collection are increased, otherwise we shall be unable to adequately tackle health inequalities for these ethnic minority patients.

ACKNOWLEDGEMENTS

We would like to thank Diane Clay, Information Officer for the Centre for Evidence in Ethnicity, Health and Diversity (CEEHD), who carried out the electronic searches, and the Advisory Board for their helpful comments on the interim study results. This work was commissioned by Cancer Research UK, and we would like to thank Vanessa Gordon-Dseagu, Catherine Foot and Ruth Yates for their support throughout this project.

REFERENCES

- Aspinall P (2009) The future of ethnicity classifications. *Journal of Ethnic and Migration Studies* 35:141735.
- Baker DW, Cameron KA, Feinglass J *et al* (2005) Patients' attitudes toward health care providers collecting information about their race and ethnicity. *Journal of General Internal Medicine* 20:895900.
- Baker DW, Hasnain-Wynia R, Kandula NR *et al* (2007) Attitudes toward health care providers, collecting information about patients' race, ethnicity, and language. *Medical Care* 45:103442.
- Buescher PA, Gizlice Z and Jones-Vessey KA (2005) Discrepancies between published data on racial classification and self-reported race: evidence from the 2002 North Carolina live birth records. *Public Health Reports* 120: 3938.
- Cancer Research UK (2006) *PROCEED*; www.info.cancerresearchuk.org/proceed
- Centers for Disease Control and Prevention (1999) Reporting race and ethnicity data National Electronic Telecommunications System for Surveillance, 1994-1997. *MMWR Morbidity and Mortality Weekly Report* 48:30512.
- Chattar-Cora D, Onime GD, Valentine IS *et al* (2000) Colorectal cancer in a multi-ethnic urban group: its anatomical and age profile. *International Surgery* 85:13742.
- Commission for Racial Equality (2002) *Ethnic Monitoring: a guide for public authorities*; www.equalityhumanrights.com/uploaded_files/PSD/3_the_duty_to_promote_race_equality.pdf
- Commission for Racial Equality (2008) *Statistics: health care services*. London: Commission for Racial Equality.
- Department of Health (2001) *Collecting Ethnic Category Data: guidance and training material for implementation of the new ethnic categories*. London: Department of Health.
- Department of Health (2005) *A Practical Guide to Ethnic Monitoring in the NHS and Social Care*; www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4116839
- Department of Health (2007) *Cancer Reform Strategy*; www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_081006
- Farooq S and Coleman MP (2005) Breast cancer survival in South Asian women in England and Wales. *Journal of Epidemiology and Community Health* 59:4026.
- Fischbacher C, Bhopal R, Povey C *et al* (2007) Record linked retrospective cohort study of 4.6 million people exploring ethnic variations in disease: myocardial infarction in South Asians. *BMC Public Health* 7:142.
- Ford ME and Kelly PA (2005) Conceptualizing and categorizing race and ethnicity in health services research. *Health Services Research* 40:165875.
- Frost F, Tollestrup K, Ross A *et al* (1994) Correctness of racial coding of American Indians and Alaska Natives on the Washington State death certificate. *American Journal of Preventive Medicine* 10:2904.
- Gardi M (2003) *Ethnicity Data Protocols Training Presentation*. Ministry of Health, Manatu Hauora; [www.moh.govt.nz/moh.nsf/pagesns/228/\\$File/ethnicitytraining.ppt](http://www.moh.govt.nz/moh.nsf/pagesns/228/$File/ethnicitytraining.ppt)
- Gill PS, Kai J, Bhopal RS *et al* (2007) Black and minority ethnic groups. In: Stevens A, Raftery J and Mant J (eds) *Healthcare Needs Assessment: the epidemiologically based needs assessment reviews*. Oxford: Radcliffe Publishing. pp. 22739.
- Gotay CC and Holup J (2004) Ethnic identities and lifestyles in a multi-ethnic cancer patient population. *Pacific Health Dialog* 11:1918.
- Greater London Authority (2005) *London: the world in a city. An analysis of 2001 Census results*. London: Data Management and Analysis Group, Greater London Authority.
- Hasnain-Wynia R and Baker DW (2006) Obtaining data on patient race, ethnicity, and primary language in health care organizations: current challenges and proposed solutions. *Health Services Research* 41:150118.
- Hasnain-Wynia R, Pierce D and Pittman MA (2004) *Who, When, and How: the current state of race, ethnicity and primary language data collection in hospitals*. Chicago, IL: Health Research and Educational Trust and the Commonwealth Fund.
- Health Research and Educational Trust (2007) *HRET Disparities Toolkit: a toolkit for collecting race, ethnicity and primary language information for patients*. Chicago, IL: Health Research and Educational Trust.
- Health Scotland (2005) *Ethnic Monitoring Tool*; www.isd.scotland.org/isd/files/ETHNIC%20MONITORING%20TOOL.pdf

- Hospital Episode Statistics online (2004) *How Good is HES Ethnicity Coding and Where Do the Problems Lie?*; www.hesonline.nhs.uk
- Information Services Division Scotland (2009) *Improving Ethnic Data Collection for Equality and Diversity Monitoring*; www.isdscotland.org/isd/5826.html
- Iqbal G, Gumber A, Szczepura A et al (2008) *Improving Ethnic Data Collection for Statistics of Cancer Incidence, Management, Mortality and Survival in the UK*. National Library for Health: Ethnicity and Health Specialist Library.
- Jack RH, Linklater KM, Hofman D et al (2006) Ethnicity coding in a regional cancer registry and in Hospital Episode Statistics. *BMC Public Health* 6:281.
- Johnson MRD (1998) Ethnic monitoring: bureaucratic construction of a minority entity or identity. In: Williams C, Soydan H and Johnson M (eds) *Social Work and Minorities*. London: Routledge. pp. 7990.
- Johnson MRD (2001) Ethnic monitoring and nursing. In: Culley L and Dyson S (eds) *Ethnicity and Nursing Practice*. Basingstoke: Palgrave Macmillan. pp. 91106.
- Johnson MRD (2006) Ethnicity. In: Killoran A, Swann C and Kelly M (eds) *Public Health Evidence: changing the health of the public*. Oxford: Oxford University Press.
- Jones M and Kai J (2007) Capturing ethnicity data in primary care: challenges and feasibility in a diverse metropolitan population. *Diversity in Health and Social Care* 4:21120.
- Kelly JJ, Chu SY, Diaz T et al (1996) Race/ethnicity misclassification of persons reported with AIDS. The AIDS Mortality Project Group and the Supplement to HIV/AIDS Surveillance Project Group. *Ethnicity and Health* 1:8794.
- Kumarapeli P, Stepaniuk R, de Lusignan S et al (2006) Ethnicity recording in general practice computer systems. *Journal of Public Health* 28:2837.
- Lin SS, O'Malley CD and Lui SW (2001) Factors associated with missing birthplace information in a population-based cancer registry. *Ethnicity and Disease* 11:598605.
- Liverpool John Moores University (2000) *Ethnicity Profiling in Primary Care: the Princes Park Health Centre model*; www.nwph.net/nwpho/Publications/Ethnic%20HEALTH%20RE-lay.pdf
- London Health Observatory (2003) *Diversity Counts. Ethnic health intelligence in London: the story so far*; www.lho.org.uk/Download/Public/7736/1/Ethnic_Health_Intelligence_Report_4.pdf
- Office for National Statistics (2001) *UK Census 2001*; www.statistics.gov.uk/census2001/census2001.asp
- Office for National Statistics (2003) *Ethnic Group Statistics: a guide for the collection and classification of ethnicity data*; www.statistics.gov.uk/about/ethnic_group_statistics/downloads/ethnic_group_statistics.pdf
- Olatokunbo S and Bhopal R (2000) 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* 2:1068.
- Pringle M and Rothera I (1996) Practicality of recording patient ethnicity in general practice: descriptive intervention study and attitude survey. *BMJ* 312: 108082.
- Race for Health (2006) *Lambeth PCT Peer Review*; www.raceforhealth.org/storage/files/Lambeth_Outcome.pdf
- Race for Health (2007) *Ethnic Monitoring in General Practice: communicating the benefits*. Liverpool: Race for Health.
- Raleigh V (2008) Collection of data on ethnic origin in England. *BMJ* 337:a1107.
- Regenstein M and Sickler D (2006) *Race, Ethnicity, and Language of Patients: hospital practices regarding collection of information to address disparities in health care*. Washington, DC: National Public Health and Hospital Institute.
- Sheth T, Nargundkar M, Chagani K et al (1997) Classifying ethnicity utilizing the Canadian Mortality Data Base. *Ethnicity and Health* 2:28795.
- Smith LK, Botha JL, Hollingworth J et al (1999) Incidence of cancer in Leicestershire: the effects of ethnicity and deprivation. In: *Proceedings of the International Association for Critical Realism*.
- Smith LK, Botha JL, Benghiat A et al (2003) Latest trends in cancer incidence among UK south Asians in Leicester. *British Journal of Cancer* 89:7073.
- Sugarman JR and Lawson L (1993) 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* 21:3836.
- Swallen KC, West DW, Stewart SL et al (1997) Predictors of misclassification of Hispanic ethnicity in a population-based cancer registry. *Annals of Epidemiology* 7:2006.
- Townsend P, Davidson N and Whitehead M (1988) *Inequalities in Health: the Black Report and the health divide*. Harmondsworth: Penguin Books.
- United Nations Statistics Division (2003) *Ethnicity: a review of data collection and dissemination*. New York: Social and Housing Statistics Section, Demographic and Social Statistics Branch, United Nations Statistics Division.
- Warnakulasuriya KA, Johnson NW, Linklater KM et al (1999) Cancer of mouth, pharynx and nasopharynx in Asian and Chinese immigrants resident in Thames regions. *Oral Oncology* 35:4715.
- Weinick RM, Caglia JM, Friedman E et al (2007) Measuring racial and ethnic health care disparities in Massachusetts. *Health Affairs* 26:1293302.
- White A (2002) *Social Focus in Brief: ethnicity*; www.statistics.gov.uk/downloads/theme_social/social_focus_in_brief/ethnicity/ethnicity.pdf
- Woods LM, Rachet B and Coleman MP (2006) Origins of socio-economic inequalities in cancer survival: a review. *Annals of Oncology* 17:519

ETHICAL APPROVAL

This study was approved by South Birmingham Research Ethics Committee

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Gulnaz Iqbal, Warwick Clinical Trials Unit, Warwick
Medical School, University of Warwick, Coventry
CV4 7AL, UK. Tel: +44 (0)2476 150 178; email:
G.Iqbal@warwick.ac.uk

Received 17 July 2009

Accepted 4 October 2009

