Improving ethnicity data collection and ethnic minority participation in Randomised Clinical Trials

by

Gulnaz Iqbal

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Health Sciences

University of Warwick, Warwick Medical School

November 2013

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to acknowledge my supervisors Professors Janet Dunn and Margaret Thorogood for their support, guidance and encouragement throughout this journey, particularly during the difficult times when I felt I was running around in circles. To Janet, for her contribution to the Cancer Research UK CanEth project, granting me study leave to write up and encouraging me to 'let go'. To Margaret, for her contribution as co-reviewer on the literature review in Chapter Seven, and invaluable advice during the course of this research.

I would like to thank Cancer Research UK who commissioned the CanEth project which provided me with nine months of funding and allowed me to conduct part of the research contained within this thesis that otherwise would not have been possible. I am grateful to all members of the CanEth team, Professors Ala Szczepura, Mark Johnson, Sue Wilson and Dr Anil Gumber for their contributions during the course of the project. I'd also like to thank Diane Clay who ran searches for the first literature review as part of CanEth. I am also grateful to the focus group facilitators, participants and healthcare professionals who responded to the survey for their contributions.

A special mention must go to Professor Annie Young for her support, encouragement and positivity, and for use of the datasets for this research. I am extremely grateful to the local teaching hospital and the cancer research network

for making data available to me for use in this thesis. Particularly to the cancer research network lead, managers, research nurses and patients without whom Chapters Eight and Nine would not have been possible. I would also like to thank the cancer patients who took the time to complete and return questionnaires.

I am grateful to my colleagues at the University of Warwick who have supported me throughout, especially Dr Louise Hiller who stepped in as emergency proof reader and Dr Andrea Marshall who provided a sounding board during my many periods of frustration.

An especially big thank you must go to my family for their unconditional love, support and patience throughout this journey. To my wonderful husband Javid Iqbal who always believed in me and kept me going when I felt defeated. To my lovely sister-in-law Sobia Iqbal for giving up her weekends and driving 160 miles for months to babysit so I could study. To my beautiful baby boy Zain Rayhan, who I have missed dearly and with whom I cannot wait to spend some quality time.

Dedication

This thesis is dedicated to my Mum and Dad who took a leap of faith all those years ago and migrated to the UK in search of a better life. If it was not for your courage, I would not be here writing this thesis today.

Declaration and inclusion of published work

I declare this thesis is my own work except where it contains work based on collaborative research. The nature and extent of my contribution have been indicated where this is applicable. I declare this thesis has not been submitted for a degree at another university.

The work contained in Chapters Four, Five and Six formed part of an unpublished report written for Cancer Research UK who funded the CanEth project (1). The systematic literature review reported in Chapter Four, the focus groups reported in Chapter Five and the healthcare professional survey contained in Chapter Six have also been published in peer reviewed journals (2,3 and 4 respectively):

1. Iqbal G, Gumber A, Szczepura A, Johnson MRD, Wilson S, Dunn J. CanEth: Improving ethnicity data collection for statistics of cancer incidence, prevalence and survival in the United Kingdom.
2. Iqbal G, Gumber A, Johnson MRD, Szczepura A, Wilson S, Dunn JA. Improving ethnicity data collection for health statistics in the UK. Diversity in Health and Care 2009:6(4):267-285.
3. Iqbal G, Johnson MRD, Szczepura A, Wilson S, Gumber A, Dunn JA. UK ethnicity data collection for healthcare statistics: The South Asian perspective. Biomed Central Public Health 2012. DOI 10.1186/1471-2458-12-243.

4. Iqbal G, Johnson MRD, Szczepura A, Gumber A, Wilson S, Dunn JA. Ethnicity data collection in the UK: The healthcare professional's perspective. Diversity and Equality in Health and Care 2013:10:281-290.

In addition to the above report and publications, I have also presented my research at the following conferences:

1. **Iqbal G**, Gumber A, Szczepura A, Johnson MRD, Wilson S, Dunn J. The changing age structure of the UK ethnic population, 1991 to 2001. Birmingham: NCIN Cancer Conference, 2010.
2. **Iqbal G**, Gumber A, Szczepura A, Johnson MRD, Wilson S, Dunn J. CanEth: Improving ethnicity data collection for statistics of cancer incidence, prevalence and survival in the United Kingdom – public and professionals perceptions. Birmingham: NCIN Cancer Conference, 2009.
3. **Begum G**, Gumber A, Szczepura A, Johnson MRD, Wilson S, Dunn J. CanEth: Improving ethnicity data collection for statistics of cancer incidence, prevalence and survival in the United Kingdom – public and professionals perceptions. Birmingham: NCRI Cancer Conference, 2008; C4.
4. **Begum G**, Gumber A, Johnson MRD, Szczepura A, Wilson S, Dunn JA. The CanEth project: improving ethnicity data collection for statistics of cancer incidence, prevalence and survival in the United Kingdom. Clinical Trials 2007.

Abstract

UK ethnic minority patients are reported to be under-represented in clinical trials. Under-representation of any patient group within an Randomised Clinical Trial can bias trial results and subsequent extrapolation into the general population. However, the true extent of ethnic minority representation in RCTs is masked by the poor quality of ethnicity data. This thesis investigates ethnicity data collection in healthcare and the representation of ethnic minority patients in RCTs in the UK.

A systematic literature review of ethnicity data collection identified a paucity of published evidence. Self-reported ethnicity was recognised as the optimal method of data collection but training is needed to raise awareness of the importance of such data. Reasons for the gaps in ethnicity data were explored through a healthcare professional survey and focus groups with lay South Asian volunteers. The majority of healthcare professionals agreed it was important to collect ethnicity and emphasised the need for training. The focus groups revealed a willingness to provide these data, subject to being given information regarding their use.

A second systematic literature review of interventions to improve recruitment and retention of ethnic minorities to RCTs found a dearth of literature from the UK. US studies reported financial incentives, government grants and the involvement of community representatives to be effective.

Census data, hospital episode statistics data, clinical trials recruitment and reasons for non-participation, collected in one cancer research network, were used to assess the local representation of ethnic minorities within RCTs. The results did not show ethnic minorities to be under-represented, but there was insufficient evidence to rule out under-representation as a problem.

Reducing inequalities in participation in clinical trials is dependent upon having accurate and complete ethnicity data. A change in policy to mandate the collection of ethnicity data in primary care and linked through to other healthcare service providers is required.

List of Abbreviations

ACCORN	Academic Clinical Oncology and Radiobiology Research Network
BAME	Black, Asian and Minority Ethnic
BME	Black minority and ethnic
BMEG	Black minority and ethnic group
CDC	Centres for Disease Control
CDSR	Cochrane Database of Systematic Reviews
CEEHD	Centre for Evidence in Ethnicity Health and Diversity
CLRN	Comprehensive Local Research Network
CRD	Centre for Reviews and Dissemination
CRE	Commission for Racial Equality
CRN	Cancer Research Network
CRUK	Cancer Research UK
DARE	Database of Abstracts of Reviews of Effects
DoH	Department of Health
ECC	Ethics and Confidentiality Committee
HCP	Healthcare Professionals
HES	Hospital Episode Statistics
HRET	Health Research and Educational Trust
ICD-10	International Classification of Diseases (version 10)
IPRP	Individual Patient Registration Profile
MDT	Multi-Disciplinary Team

MeSH	Medical Subject Headings
NCIN	National Cancer Intelligence Network
NCRN	National Cancer Research Network
NIH	National Institute of Health
NMSC	Non-Melanoma Skin Cancer
ONS	Office for National Statistics
PCT	Primary Care Trust
PICOS	Population, Interventions, Comparators, Outcomes & Study designs
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QOF	Quality Outcomes Framework
R&D	Research and Development
RCT	Randomised Controlled Trial
SEER	Surveillance, Epidemiology and End Results
SLEH	Specialist Library for Ethnicity and Health

Chapter 1 Introduction

1.1 Rationale for research topic

Participation of ethnic minorities has been reported to be under-represented in clinical trials (Hussain-Gambles 2004, Godden et al, 2010). As a South Asian working in cancer clinical trials for over a decade, I was aware of evidence suggesting patients entering clinical trials had better outcomes even when randomised to standard care, also referred to as a 'trial effect' (Karjalainen and Palva, 1989, Braunholtz et al, 2001). It is suggested that patients benefit from active follow-up and closer monitoring and management of their disease. As the daughter of parents speaking little English, I played a pivotal role when my father was invited to participate in a phase III diabetes Randomised Controlled Trial (RCT). Patients like my father who do not have an English speaking family member to accompany them to appointments and act as an interpreter are still likely to miss the opportunity to participate in a trial.

The true extent of ethnic minority representation in RCTs is masked by the poor quality of ethnicity data in healthcare. Through this thesis, I investigate ethnicity data collection in healthcare, review the literature using interventions to improve ethnic minority participation in RCTs, investigate the representativeness of ethnic minorities in RCTs and explore barriers to participation.

I started my career as a statistical reviewer on a systematic literature review looking at barriers and attitudes to participation in clinical trials from the patient and healthcare professional perspectives (Ward et al, 2000). The review revealed the patients participating in clinical trials to be mainly White, English speaking, middle-class men. Barriers to trial participation identified included: distrust of healthcare professionals, lack of time and resources and language issues. Language was important as most patient information sheets and consent forms were only available in English. This automatically excluded all non-English speakers from clinical trials unless interpreters were readily available. However, patient information sheets were translated into two languages (Urdu and Gujarati) in a cancer trial using prophylactic anti-coagulants for which I was the statistician. Unfortunately, they did not improve uptake which may have been due to poor staff preparedness or poor literacy levels in certain ethnic minorities (Department of Health, 1999, Szczepura et al, 2005).

I went on to analyse a West Midlands Occupational Urothelial Tumour Unit study exploring survival and delay times in the diagnosis and treatment of bladder cancer (from onset of first symptoms to treatment). The results of the study showed delays affected overall survival, with shorter GP and hospital delays resulting in poorer survival, this was as expected, with patients with severe symptoms rushed through quickly (Wallace et al, 2002). It was suspected that delay times may vary by geographical area, otherwise known as the 'postcode lottery'. All patients with a valid postcode were assigned a Townsend Material Deprivation Score. The analysis

confirmed material deprivation as a significant prognostic variable along with other patient and tumour characteristics. However, the analysis also showed that delay times did not significantly vary across geographical area (Begum et al, 2004). Feedback from presentation of the results suggested ethnicity may be a contributing factor believing it to be a proxy for deprivation and vice versa (Begum et al, 2002). Unfortunately, ethnicity data were not available for this study. However, it was around this time that I first developed an interest in inequalities in cancer.

Subsequently, my speciality became the analysis and reporting of haematological trials, which included the Medical Research Council Multiple Myeloma trials and studies of leukaemia. I also became aware of the increased risk of developing multiple myeloma in the African and Caribbean communities (Smith et al, 2005). I was then involved in a study examining the occurrence of acute Graft versus Host Disease and transplant-related mortality by ethnic group in patients undergoing standard sibling allogeneic stem cell transplantation. Indications of an increased risk in the non-Caucasian group were found which may have been attributable to differences in tumour biology, socioeconomic factors like nutrition, or post-transplant care (Karanth et al, 2006).

Given the paucity of ethnicity data and the heavy consequences of not identifying and facilitating patients from ethnic minority groups into clinical trials described above, the research topic for this thesis was formed.

The thesis investigates ethnicity data collection and the issue of ethnic minority under-representation in RCTs, the consequences of which are clouded by the poor collection of ethnicity data in healthcare. This research was started by a nine month grant from Cancer Research UK (CRUK) who commissioned the Cancer Ethnicity (CanEth) feasibility study to identify methods of collecting accurate data on ethnicity and recognised this as being essential to inform policy makers, funders and public health experts on the cancer incidence, prevalence and outcomes.

1.2 Ethnic minorities in the UK

1.2.1 UK migration

The British Nationality Act (1948) passed due to a shortage of labourers, allowed people born in British territory to enter and work in the UK without the need for a visa. The 22nd June 1948 saw the first arrivals from the West Indies and marked the start of mass immigration into the UK, until 1972 when new legislation restricting immigration to those with work permits or with family born in the UK was introduced. Immigration from the former British Empire colonies of India, Pakistan, Bangladesh, the Caribbean, South Africa, Kenya and Hong Kong was vast, and saw the ethnic minority population of the UK rise from a few thousand in 1948 to 1.4 million in 1970. Further to this, political changes in Africa led to a stream of migration of African Indians between the 1960s and early 1970s (Gill et al, 2007). Later immigrants originated from Pakistan and Bangladesh or were the families of the earlier arrivals. In recent years, we have seen the arrival of new migrants from

Eastern Europe following the enlargement of the European Union in 2004. Numbers of asylum seekers and refugees of Somali and Middle-Eastern origin have also increased (The Migration Observatory, 2013).

1.3 Ethnicity data collection

1.3.1 Census

Ethnicity data collection first commenced in the UK in the form of country of birth collected in the census from 1841 to 1991. In 1991, the ethnicity question was introduced providing ten categories based upon immigration routes. This was expanded to 16 categories in 2001 and 18 for the 2011 census. These changes were accompanied by free text boxes to capture free text. The expansions were a result of large numbers of the population selecting 'other' and entering free text because they did not identify with the choice of categories available, and also to capture the growing mixed population.

1.3.2 Health

Ethnicity data collection was also initiated in other areas besides the census. It became a requirement to collect ethnicity data as part of the Hospital Episode Statistics (HES) data in 1995 but was poorly implemented. The levels of missing data and invalid codes rendered the data unusable (Aspinall, 2000). However, some improvement has been reported in Finished Consultant Episodes with a reduction in 'not known' and 'not stated' codes from 23.9% in 2004/05 to 8.6% in 2009/10

(HES online, 2011). In 2004, a single Quality and Outcomes Framework (QOF) point out of a possible thousand was offered for collection of ethnicity data for all new registrations in primary care, and again this was a weak incentive which did not have much impact. The system could also be cheated by selecting 'other', or 'not known' which would mean the field was complete but did not provide any useful information, the incentive was recognised to be ineffective and abandoned (Johnson, 2012).

1.4 Ethnicity

The term 'ethnicity' is derived from the Greek word 'Ethnos' meaning 'tribe of people'. Ethnicity is a multi-dimensional and subjective concept, which means that self-reporting is the most accurate method of capturing this information. Ethnic groups can be broadly defined as people from the same geographical location or those who have a common language, religion or culture. There is no hard or fast definition of ethnicity and the area is subject to great debate with researchers using a diverse range of terminology, which has meant that publications relating to ethnic minorities are not easy to find. Bhopal (2004) attempted to define the concept and measures of ethnicity in the form of a glossary style article with the hope that it might lead to an internationally applicable glossary and provide a basis for indexing terms. The essential components of ethnicity can be said to comprise of the following items but this list is by no means exhaustive:

- Country of birth

- Country of origin
- Parents country of origin
- Religion
- Language
- Diet
- Nationality
- Race
- Culture
- Dress

Race is an out of date term and is considered unacceptable in the UK (Gill et al, 2007). Historically, race was used prior to ethnicity and is still collected in some countries such as the USA today. Race refers to physical attributes such as, skin colour, facial features, hair colour and texture. Skin colour is an inaccurate indicator of ethnicity, it is not possible to distinguish between white groups such as English, Scottish, Welsh or other Europeans nor between some non-whites, e.g. Indian, Pakistani or Bangladeshi. The items considered to be of national importance can first and foremost be found in the census.

1.4.1 Ethnicity indicators

1.4.1.1 Ethnic group

Prior to 1991, country of birth was the only indicator of ethnicity collected. Ethnicity was first incorporated in the 1991 census as a ten category item (Table 1.1). The categories followed immigration roots and combined geographical origin with skin colour. The intention was to increase the accuracy of ethnic group classification

where subjects originated from a multi-cultural continent, e.g. Black African, Black Caribbean. In 2001, the census categories were supplemented with the addition of a new 'Mixed' categories, expansion of the 'White' category and the introduction of an 'Other' category, with space for free text in each category. The categories were arranged into five broad groups: White, Black, Mixed, Asian, Chinese and Other.

The 2011 census saw the introduction of two further classifications, 'Gypsy or Irish traveller' and 'Arab' accumulating in 18 categories. Other changes included relocating the Chinese category from 'Other ethnic group' in 2001 to the 'Asian/Asian British' category. A 'not stated' code exists for coding purposes to be used by those collecting the data but not offered on forms or verbally in order to minimise missing data (Table 1.1). No subsequent changes have been made to ethnic minority categorisation in the UK census.

Table 1.1 Ethnicity categories in thirty years of the England and Wales census 1991-2011

1991		2001		2011	
0	White	A	White	A	White
1	Black Caribbean	1	British	1	English/Welsh/Scottish/Northern Irish/British
2	Black African	2	Irish	2	Irish
3	Black other	3	Any other white background	3	Gypsy or Irish Traveller
4	Indian	B	Mixed	4	Any other White background
5	Pakistani	4	White and Black Caribbean	B	Mixed/multiple ethnic groups
6	Bangladeshi	5	White and Black African	5	White and Black Caribbean
7	Chinese	6	White and Asian	6	White and Black African
8	Other Asian	7	Any other mixed background	7	White and Asian
9	Any other ethnic	C	Asian or British Asian	8	Any other Mixed/multiple ethnic background
	Not stated	8	Indian	C	Asian/British Asian
		9	Pakistani	9	Indian
		10	Bangladeshi	10	Pakistani
		11	Any other Asian background	11	Bangladeshi
		D	Black or Black British	12	Chinese
		12	Caribbean	13	Any other Asian background
		13	African	D	Black/African/Caribbean/Black British
		14	Any other Black background	14	African
		E	Other Ethnic Groups	15	Caribbean
		15	Chinese	16	Any other Black/African/Caribbean background
		16	Any other ethnic group	E	Other ethnic group
			Not stated	17	Arab
				18	Any other ethnic group
					Not stated

1.4.1.2 Religion

Religion was first introduced as a census question in 2001 and remained unchanged for the 2011 round, it continues to be optional. The question comprises of seven responses: 1) Christian, 2) Buddhist, 3) Hindu, 4) Jewish, 5) Muslim, 6) Sikh, and 7) any other religion. The Office of National Statistics also distinguished between degree of religious practice (active faith and participation in religious activity and worship) and religious identity (belonging to a religious community even if the religion is not practiced). As previously, the 'not stated' option was not offered either verbally or as a category on a form.

1.4.1.3 Language

Language was collected as part of the 2011 census in two parts. The first asked for language to be specified if it was not English, with the second part asking respondents to indicate their level of proficiency in English using a Likert scale.

1.4.1.4 Country of birth

Country of birth is routinely recorded on UK birth and death certificates and has been collected in the census since 1841 this has often been used as a proxy for ethnicity mainly because of its availability. However, it has two main weaknesses. Firstly, it is not possible to distinguish immigrants from the children of British residents born overseas, even though these individuals would identify themselves as White. Secondly, it is unable to identify the British born offspring of immigrants

and with over 50% of ethnic minorities in 2001 born in the UK, country of birth is no longer a reliable indicator of ethnicity.

1.4.1.5 Nationality

Nationality was collected as part of the census between 1841 and 1961 before being abandoned. It was first collected in order to identify those of Scottish and Irish origin. However, following the mass inward migration of the 1960s the question caused confusion and data were difficult to collect. However, nationality and feeling like a member of the society in which you live are important parts of identity and have been recognised by the Office of National Statistics. The 2011 census once again included a question about national identity, allowing individuals to choose from six categories: English, Welsh, Scottish, Northern Irish, British or Other, with space provided to capture free text.

1.4.2 Limitations of common reporting categories

Data from the census surveys are often banded together to create two broad categories referred to as the 'White' and 'non-White' or the more recently established 'Black Minority and Ethnic Groups (BMEGs)' and 'Black, Asian and Minority Ethnic (BAME)'. Grouping data in this way means important differences between groups may be missed. For example, people of Irish, Polish, and European origin, would be lost amongst the majority White British population. These groups are known to have similar problems with health inequalities and inequities in access to healthcare as the non-White population and the need to identify these

communities is of equal importance (Aspinall, 1998). In other reports, the term 'White' is used inconsistently and varies between contexts, sometimes it is used as above, referring to all people identifying themselves as 'White', but for a significant proportion it has been used to describe people of 'White British' origin only (Department of Health, 2009).

The term 'Black' is often used without clarification of whether this refers to African or Caribbean communities. These communities differ from one another in terms of their migration history, religious tradition, culture and language (King's Fund, 2006). African-Caribbean was abbreviated to Afro-Caribbean, however, this is no longer deemed acceptable and is considered to be an offensive term relating to hair texture (Department of Health, 2009).

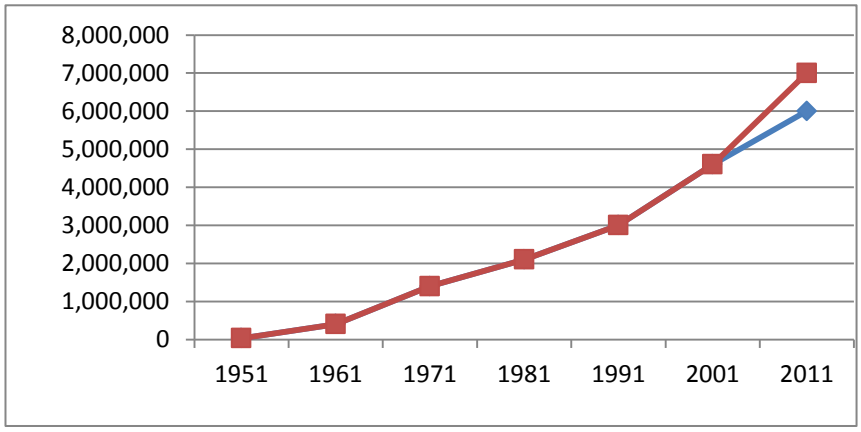
Bi-cultural terms such as 'Asian' or 'Asian British' are now used in many contexts including the British census (Aspinall, 2003). However, it is not clear whether 'Asian' refers to people originating from South Asia or East Asia. In the UK, most would understand the use of the word to indicate someone originating from the Indian subcontinent. However, in the USA and Canada this term is used to indicate someone of East Asian origin. Caution needs to be exercised as members of the 'Asian' communities vary dramatically in terms of their migration history, country of origin, education, social class, religion, culture and language.

1.4.3 The changing composition of the UK population

In order to get a clearer picture of the UK population, the most accurate source of data were population surveys such as the census. The last census conducted in the UK was in 2011. However, as not all data had been released at the time of this write-up, I used data from both 2001 and 2011 to construct a picture of the ethnic minority population.

The rapid growth of the non-white population since the passing of the British Nationality Act can be seen in Figure 1.1. The population figures for 1951 and 1961 are estimates only, whilst the figures for 1971 and 1981 are based on country of birth data. An increase of 53% was observed between the non-White populations of 2001 compared to 3 million in 1991. Calculations based upon UK population growth patterns predicted the non-white population would be between 6 and 7 million by the 2011 census (Commission for Racial Equality, 2007). This was exceeded with the non-White population reported to be over 7.5 million in England and Wales.

Figure 1.1: Growth of the Non-White population of Great Britain



The 2001 census reported 6.5 million people in England and Wales identifying themselves as belonging to an ethnic minority group, this is inclusive of 4.5 million (8.7%) classified as belonging to a 'non-white' ethnic group (Office for National Statistics, 2001, Table 1.2). Over a third of all minorities were either 'Asian' or 'British Asian', an additional 17.5% as Black or Black British (Table 1.3). Other Whites were the largest group in 2001 accounting for 20.7% of the total minority ethnic population, followed by Indians at 15.9% and the Pakistanis group at 11.0%.

Table 1.2 also shows over 661,034 people in 2001 identifying themselves as of 'mixed' origin and a further 219,754 who selected 'other', most probably because the existing choices did not fit their perception of their own identity. Data from the 'other' categories of the 2001 census have shown approximately 4% of the population were unable or unhappy to choose their ethnic group from the available categories. 'Other White' was the largest of these groups, and 80% of the 1.3 million people choosing this category were born overseas.

The most common ‘any other ethnic group’ entries were Arab (n=45027), Filipino (n=37590) and Middle Eastern (excluding Israeli, Iranian and Arab, n=37397) (Gardener and Connolly, 2005). White and Black Caribbean was the largest mixed ethnic group in 2001 (n=237,000) followed by White and Asian (n=189,000) and other mixed (n=155,000); White and Black Africans was the smallest (n=78,000) (Bradford, 2006). In total, there were over 80 additional entries suggesting that the 16 census categories are inadequate to deal with the multi-ethnic fabric of the UK and more open-ended questions should be used (Aspinall, 2009).

In comparison, data from the 2011 census showed the proportion of non-white minorities had increased to 14.1%, with an additional 4.4% identifying as Other White (Table 1.3). Indians remained the largest non-White group accounting for 2.5%; however, the mixed group was the second largest group (2.2% in total) overtaking Pakistanis (2%).

The ethnic minority landscape in the UK has markedly changed over the last two decades. The impact of this changing landscape on patients suffering with cancer will be explored.

Table 1.2 The England and Wales population by ethnic group, Census 2001 and 2011

2001	n	%		2011	n	%
White	47,520,866	91.3		White	48,209,395	86.0
British	45,533,741	87.5		English/Welsh/Scottish/Northern Irish/British	45,134,686	80.5
Irish	641,804	1.2		Irish	531,087	0.9
				Gypsy or Irish Traveller	57,680	0.1
Any other white background	1,345,321	2.6		Any other White background	2,485,942	4.4
Mixed	661,034	1.3		Mixed/multiple ethnic groups	1,224,400	2.2
White and Black Caribbean	237,420	0.5		White and Black Caribbean	426,715	0.8
White and Black African	78,911	0.2		White and Black African	165,974	0.3
White and Asian	189,015	0.4		White and Asian	341,727	0.6
Any other mixed background	155,688	0.3		Any other Mixed/multiple ethnic background	289,984	0.5
Asian or British Asian	2,500,685	4.8		Asian/British Asian	4,213,531	7.5
Indian	1,036,807	2.0		Indian	1,412,958	2.5
Pakistani	714,826	1.4		Pakistani	1,124,511	2.0
Bangladeshi	280,830	0.5		Bangladeshi	447,201	0.8
Chinese	226,948	0.4		Chinese	393,141	0.7
Any other Asian background	241,274	0.5		Any other Asian background	835,720	1.5
Black or Black British	1,139,577	2.2		Black/African/Caribbean/Black British	1,864,890	3.3
African	479,665	0.9		African	989,628	1.8
Caribbean	563,843	1.1		Caribbean	594,825	1.1
Any other Black background	96,069	0.2		Any other Black/African/Caribbean	280,437	0.5
Other Ethnic Groups	219,754	0.4		Other ethnic group	563,696	1.0
				Arab	230,600	0.4
Any other ethnic group	219,754	0.4		Any other ethnic group	333,096	0.6
Non-White population	4,521,050	8.7		Non-White population	7,866,517	14.0
All ethnic minorities	6,508,175	12.5		All ethnic minorities	10,941,226	19.5
Total	52,041,916			Total	56,075,912	

Table 1.3 The England and Wales ethnic minority population, Census 2001 and 2011

2001	n	%		2011	n	%
White	1,987,125	30.5		White	3,074,709	28.1
British	45,533,741	n/a		English/Welsh/Scottish/Northern Irish/British	45,134,686	n/a
Irish	641,804	9.9		Irish	531,087	4.9
				Gypsy or Irish Traveller	57,680	0.5
Any other white background	1,345,321	20.7		Any other White background	2,485,942	22.7
Mixed		10.2		Mixed/multiple ethnic groups	1,224,400	11.2
White and Black Caribbean	237,420	3.6		White and Black Caribbean	426,715	3.9
White and Black African	78,911	1.2		White and Black African	165,974	1.5
White and Asian	189,015	2.9		White and Asian	341,727	3.1
Any other mixed background	155,688	2.4		Any other Mixed/multiple ethnic background	289,984	2.7
Asian or British Asian		38.4		Asian/British Asian	4,213,531	38.5
Indian	1,036,807	15.9		Indian	1,412,958	12.9
Pakistani	714,826	11.0		Pakistani	1,124,511	10.3
Bangladeshi	280,830	4.3		Bangladeshi	447,201	4.1
Chinese	226,948	3.5		Chinese	393,141	3.6
Any other Asian background	241,274	3.7		Any other Asian background	835,720	7.6
Black or Black British		17.5		Black/African/Caribbean/Black British	1,864,890	17.0
African	479,665	7.4		African	989,628	9.0
Caribbean	563,843	8.7		Caribbean	594,825	5.4
Any other Black background	96,069	1.5		Any other Black/African/Caribbean background	280,437	2.6
Other Ethnic Groups		3.4		Other ethnic group	563,696	5.2
				Arab	230,600	2.1
Any other ethnic group	219,754	3.4		Any other ethnic group	333,096	3.0
All ethnic minorities	6,508,175			All ethnic minorities	10,941,226	

1.5 The cancer burden

According to 'Improving Outcomes: a strategy for cancer', 'every year around 250,000 people in England are diagnosed with cancer and around 130,000 will die from the disease, making it the leading cause of mortality in people under the age of 75'. Currently, 1.8 million people are living with or beyond a cancer diagnosis (Department of Health, 2011).

Cancer survival has improved greatly, since the early 1970s when only 28% survived beyond five years increasing to approximately 50% in more recent years. These improvements have been reported in the majority of the common cancers. Despite these changes, there are some groups for whom there is no evidence of having benefited from these improvements; ethnic minorities being one such group. Cancer registries initiated the collection of ethnicity data within the last few years and to date have obtained information through linkage with HES. Due to incomplete data, there has been little reliable information on the patterns of cancer incidence, mortality and survival, specific to the UK's ethnic minorities. The source of evidence on inequalities stemmed from small scale studies, such as Bowen et al (2008), London Health Observatory (2004) and Farooq and Coleman (2005).

Other members of society who have not benefited from the improvements in cancer survival include those living in deprived areas, although there is speculation that the overlap between these two groups is large. However, due to the gaps in ethnicity data there is patchy evidence reporting the correlation between ethnic

origin and deprivation. The Department for Work and Pensions (2001) reported approximately 66% of the Bangladeshi community to be living in low income households, and an estimated 50% of the Black Caribbean community. These figures were much higher when compared to 21% of the White population.

Cancer risk is known to be increased by many of our lifestyles choices such as, smoking, excessive alcohol consumption, poor diet, obesity, lack of exercise and unprotected exposure to the sun (Gordon-Dseagu, 2008). People from ethnic minority backgrounds are more likely to practice 'risky behaviours', these are discussed in more detail below.

1.6 Disparities in cancer risk factors by ethnic group

1.6.1 Smoking status

Smoking is a major cause of preventable cancer deaths. Tobacco is known to cause 90% of lung cancer cases in the UK (Gordon-Dseagu, 2008). This risky behaviour has been shown to be more common in people from ethnic minority groups, for example, smoking rates were reported to be the highest in Bangladeshi males (44%) followed by Irish males (39%), compared to 27% in the general population. However, Bangladeshi women are more likely to chew tobacco (26%) than smoke cigarettes which is reported to increase the risk of head and neck cancers (White, 2002). The General Household Survey of 2005, also reported elevated smoking rates in males and females of mixed background. Rates remained highest in the Bangladeshi group at 45%, followed by males of 'White and African', and 'Other

Mixed' origin at 38% and 39%, respectively. In females, 'White and Asian' women were found to have the highest rate of smoking at 33%, second to those of 'White and Black Caribbean' origin with 29%.

The Health Survey for England 2004, which focused on the health of minority ethnic groups, also reported the use of chewing tobacco and found rates to be approximately 9% in Bangladeshi males and 16% of females (The Information Centre, 2006). Paan or 'betal leaf chew' originates from South East Asia and is commonly used as a breath freshener, palate cleanser and believed to aid digestion. There are many varieties, but usually consist of mixed spices such as cardamom and anise, lime paste, grated coconut, nuts, small piece of sweets and often tobacco. The concoction is wrapped in betal leafs and chewed but not usually swallowed. A small study by (Williams, 1999) found 78% of Bangladeshis to chew paan, with women chewing more frequently and more likely to add tobacco.

1.6.2 Diet

Diet can have an effect on the risk of cancers in the bowel, stomach, head and neck as well breast and prostate with an estimated third of deaths from cancer are related to diet (Gordon-Dseagu, 2008). Poor diet, with a high fatty and sugary content combined with a lack of exercise have seen the obesity rates soar in many countries all over the developed world. The 2004, Health Survey for England revealed many ethnic minority communities to have a higher consumption of five or more portions of fruit and vegetables a day than the general white population.

Approximately, 42% of Chinese women, 36% of Indian women and 40% of Chinese and Indian men consumed five or more portions of fruit and vegetables a day, compared to 23% of men and 27% of women in the general population. However, this was not true of all ethnic groups; consumption rates were lower than in the general population for Irish males (26%) and Bangladeshi females (28%). Further to this, the survey also found a significantly increased amount of salt used in the preparation of food but fat consumption was shown to be lower than that of the general population for all ethnic groups.

1.6.3 Alcohol consumption

A relationship between excessive alcohol consumption and cancers of the mouth, larynx, oesophagus, liver, breast and bowel has been documented, accounting for an estimated 6% of cancer deaths (Cancer Research UK, 2008). According to the Health Survey of England 2004 (The Information Centre, 2006) all ethnic minorities are less likely to drink alcohol than the general population, with the exception of the Irish community. 71% of Irish men and 53% of Irish women were more likely to exceed recommended alcohol limits. Large proportions of other ethnic groups are likely to identify themselves as non-drinkers (Gordon-Dseagu,2008).

1.6.4 Exercise

Inadequate levels of exercise have been reported to be responsible for approximately 5% of all cancer deaths (Colditz et al, 1997, Wolin et al, 2009). Increased risk associated with a lack of exercise has been shown in colon and breast

cancer, links have also been reported with cancers occurring in the womb, lung and prostate. Levels of exercise were found to be particularly low in males and females of Pakistani and Bangladeshi origin, with many taking less than 30 minutes of moderate exercise a week (The Information Centre, 2006).

1.6.5 Obesity

Lack of exercise and poor diet have contributed to the increased rates of obesity which are known to increase the risk of many diseases including cancer. Cancer Research UK (2006) stated "13,000 cases of cancer could be prevented each year in the UK if no one had a BMI greater than 25". The risk of breast cancer has been reported to increase by up to 30% in post-menopausal obese women (Lahman et al, 2004, Van Den Brandt et al, 2000). In colon cancer this risk increased by approximately 25% in men falling into the overweight category and by 50% for those in the obese category (Moghaddam et al, 2007).

According to a recent report by the National Obesity Observatory (Gatineau and Mathrani, 2011) which included data from the 'Health of Minority Ethnic Groups Survey' conducted by the Department of Health (2005) obesity rates were found to be the highest in women of Black African and Black Caribbean origin followed by women of Pakistani origin (38%, 32% and 28% respectively, compared to 23% of women from the general population. In contrast, obesity rates were lower in men of Bangladeshi and Chinese background (6% and 6% respectively) when compared to 23% of men from the general population.

1.7 Variations in cancer incidence, mortality and survival by cancer site and ethnic group

Up until the recent publication by the National Cancer Intelligence Network (NCIN) on 'Cancer incidence and survival by major ethnic group, England, 2002-2006' cancer statistics by ethnic group were unavailable (NCIN, 2009). This document reporting national figures is the first of its kind, but nevertheless is based upon incomplete data. Ethnicity was reported to be missing/unobtainable for approximately 25% of cancer registrations used in the analysis, and the findings are presented for broad ethnic categories only. Despite its limitations the data shows people of Asian, Chinese, and Mixed origins to unanimously be at lower risk of developing cancer when compared to the general White population. In spite of this, Asians were found to have higher rates across three cancer sites (liver, head and neck and cervical) but were at lower risk of developing the four most common cancers (breast, prostate, lung and colorectal).

Overall, the Black group were reported to have higher rates of prostate, stomach, liver and multiple myeloma, but had lower rates of the three common cancers (breast, lung and colorectal).

1.7.1 Cancers of the head and neck

Females of South Asian origin were shown to have a statistically significant increased rate of mouth cancer; this difference remained true after adjusting for

age. However, there was some evidence to suggest a trend in males of South Asian origin having a lower risk than in the general white population (NCIN, 2009).

1.7.2 Breast

Bowen et al (2008) reported a difference in the age and type of breast cancer in women of African and West Indian origin living in the UK. This was a small study and the validity of the findings have been questioned. The research concluded women of African and West Indian origin were diagnosed on average 21 years younger than the White British population. Furthermore, the cancer was reported to be a more aggressive form (triple negative), known to be unresponsive to conventional drugs, hence resulting in poorer survival. These women will also fall under the standard screening age bracket of 50 to 70 years. However, women with a strong family history can request screening from the age of 40. Data from the NCIN report, 2009 show survival of Black women with breast cancer aged 15-64 was reported to be 85% at three years compared to 91% in the general population. No differences were found when considering those aged 65 or over or when combining age groups.

In contrast, Farooq and Coleman (2005) who also reported differences in breast cancer by ethnic group have shown women of South Asian origin to have a lower incidence of breast cancer than the general White population. There was also some evidence to suggest survival may be different in this group. The NCIN (2009) reported three year survival from breast cancer to be significantly reduced in South

Asian women between the ages of 16-64 years, but no differences were detected in those aged 65 years and over.

According to the Breast Cancer Care Policy Briefing (2005) 32% of women of ethnic background knew little about breast cancer, 43% reported never checking their breasts for lumps, 56% of those who performed examinations reported not knowing what to look for, and only 38% considered a lump to be symptomatic of breast cancer. The briefing also reported a lack of uptake of breast screening, with 45% of women of screening age having never attended screening.

1.7.3 Prostate

Prostate cancer incidence and mortality amongst men of Black Caribbean and African origin is reported to be dramatically higher, with rates up to three times higher than that of the general population (Prostate Cancer Charter for Action, 2005). Rates are also shown to be elevated in other ethnic minority groups, however, these are to a lesser extent. The NCIN 2009, data confirmed the elevated risk in males of black origin and reported risk ratios between 1.1 and 3.4 by age group (<65 years and >=65 years respectively) compared to white males.

In the USA, the recommended age for prostate cancer screening has been lowered from the standard age of 50 to 40 years in the Black American population (BBC News, 2001). In the UK, a survey of men of African Caribbean background conducted in Birmingham found 22% of men did not know where the prostate

gland was and a further 50% did not know the function of the gland (Ethnic Minority Cancer Awareness Week, 2008).

1.7.4 Liver

The risk of liver cancer was estimated to be between 1.5 and 3 times higher in the South Asian population when compared to the general population. This difference remained true across gender, and age group (<65, ≥65) (NCIN, 2009). These data also pointed to an increased rate of liver cancer in males and females of black origin.

1.7.5 Cervical

Females of South Asian origin were found to have a significantly higher risk of cervical cancer compared to women from the general population. However, this pattern was only apparent in women of 65 years or over and was lost when combining with women <65 years (NCIN, 2009). This same pattern was observed but to a lesser extent in women of black origin aged 65 or over.

1.7.6 Other disease areas

Other diseases areas where marked differences in incidence/prevalence rates by ethnic group have been reported include diabetes, coronary heart disease, obesity, and mental illness.

There is a wealth of evidence reporting a higher incidence of non-insulin dependent diabetes in South Asians and Black Caribbean groups (Aspinall and Jacobson, 2004). This difference is also reflected in diabetes associated mortality, with patients of both Caribbean and South Asian origin 3.5 times more likely to die from causes related directly to their disease. In addition to the higher incidence, there are also reports suggesting onset of disease may also be earlier in South Asians compared to the White population.

Diabetes is also thought to be a risk factor for Coronary Heart Disease (CHD) and could explain the excess reported in South Asians. Despite the evidence showing an unequal burden of diabetes Aspinall And Jacobson (2004) stated "non-insulin dependent diabetes remains undiagnosed in up to 40% of Asian diabetics and several studies report inadequate quality of health care for Asian and African-Caribbean diabetics and poor compliance arising from patients' lack of knowledge about the disease and its management through the inappropriateness of health information".

1.8 Randomised Controlled Trials (RCTs)

1.8.1 What is a randomised controlled trial?

Clinical trials are the gold standard in medical research to test the efficacy and safety of new drugs, procedures or treatments. Many phase III clinical trials also include health economics and Quality of Life components. It is important for trials

to recruit a representative population in order to ensure external validity so the findings can be extrapolated to the population of interest.

Low participation rates in cancer trials led to the inception of the National Cancer Research Network (NCRN) in 2001. The network comprised of 32 local research networks and 700 staff. Over 330,000 cancer patient have taken part in clinical trials since its establishment.

1.8.2 Under-representation of ethnic minorities in clinical trials

There is some evidence to suggest clinical trial participants have better outcomes than non-trial participants with patients benefiting from a systematic treatment schedule as opposed to a clinician determined schedule (Karjalainen and Palva, 1989). Participating in a trial can offer closer monitoring and management of disease and extra follow-up for all trial participants. Excluding ethnic minority groups from clinical trials not only raises issues in terms of extrapolation of the results, but also in the equity of access to clinical trials. Published articles rarely report exclusion criteria but the ability to speak English had been an eligibility criteria for many trials in the past (Murray and Buller, 2007). A review of clinical trial exclusion criteria reported many trials excluded ethnic minorities without clinical or scientific justification (Britton et al, 1999). The majority of trial participants have been reported to be white middle-class men (Killien et al, 2000). This contradicts the government's NHS plan, which aimed to tackle inequalities and

provide culturally appropriate and accessible healthcare for all (Department of Health, 2005). If there is no scientific rationale for the exclusion of ethnic minorities from clinical trials this could indicate a form of discrimination or institutional racism as indicated by the Macpherson report (1999).

Events such as the Tuskagee experiment conducted between 1932-1972 and the thalidomide tragedy of the 1960s led to the creation of policies to protect vulnerable groups such as women (particularly of child bearing age) and ethnic minorities in research. It was in 1993, after concerns were raised regarding the mass exclusion of women and ethnic minorities from research and the applicability of results to these populations that resulted in a change in US government policy (National Institute for Health, 2001). The National Institute of Health Revitalisation Act 1993, requires all research funded by the USA NIH to include a representative sample of women and ethnic minorities unless there is a clinical or scientific reason for exclusion.

Under-representation of ethnic minorities in clinical trials means that results cannot be extrapolated to the whole population. It is known that pharmacodynamics, pharmacokinetics and now pharmacogenomics, can vary by ethnic group, as well as there being differences in disease characteristics (Krecic-Shepard et al, 2000, Yasuda et al, 2008, O'Donnell and Dolan, 2009). Ethnic differences have been reported in the effects of chemotherapy for several drug classes, such as antimetabolites, anthracyclines and alkylating agents (O'Donnell and Dolan, 2009).

Underlying differences such as this will continue to be of clinical importance despite the increasing acculturation of current ethnic minorities groups. Additionally, interactions between environment, genetics and culture are also known to play a part in drug metabolism (Matthews, 1995). Differences in health behaviours have also been reported to vary between ethnic groups stemming from cultural, religious and socio-economic factors making it equally important to recruit representative samples in non-drug trials (Gatineau and Mathrani, 2011).

1.8.3 Barriers to participation

Hussain-Gambles (2004) reviewed the literature and found the key motivations for participating in clinical trials were altruistic factors, benefits to own health, effective follow-up, clinician influence, communication style of doctor/nurse and satisfaction with previous experience. Barriers to patient participation included the additional demands on the patient (e.g. extra procedures and appointments, travel, childcare costs and similar), geographical location of the study site, complexity of protocol, patient having preference for a particular treatment or no treatment, listed side-effects of drugs, fear of experimentation, poor comprehension of clinical trials, distrust of hospital or medicine, process of gaining consent and socio-cultural aspects. However, barriers to participation reported by ethnic minorities differed somewhat to include:

- Mistrust
- Language
- Cultural barriers

- Importance of family
- Gender
- Health beliefs
- Modesty
- Religion
- Age
- Geographical location
- Lack of familiarity/lower awareness of trials
- Socio-economic barriers

In a review of barriers to clinical trial accrual for under-represented populations (defined as ethnic minorities, the elderly population, rural residents and those of low socioeconomic status) Ford et al (2008) categorised barriers into three groups as follows: 1) clinical trial awareness, 2) opportunities to participate and 3) the acceptance of enrolment, or in other words reasons for choosing not to participate. The results revealed the barriers relating to the opportunity to participate to be most frequently reported. Opportunity barriers included older age, ethnicity and race, no/low health insurance and exclusions based upon co-morbidities.

1.9 Addressing health inequalities

"The NHS is under a legal and moral obligation to provide services to all people who need them, regardless of the gender, age or ethnic background" (quoted from the Kings Fund: Access to healthcare and minority ethnic groups briefing, 2006). In fact, it is the duty of all public authorities, including the NHS to comply with the Race Relations Amendment Act 2000 now the Equality Act 2010. The government clearly

justifies the need for action in the Department of Health's Race Equality Scheme 2005-2008 (Department of Health, 2005a). It states "the NHS increasingly needs to take into account not only cultural and linguistic diversity but also needs to be able to cater for varying lifestyles and faiths". This suggests providing a service that is more responsive to the needs of individual patients. The scheme also identifies issues of health inequalities and inequities in access to service by ethnic group as cause for action.

Gill et al (2007) in his needs assessment of ethnic minority groups says: "In the past, data on minorities groups have been presented to highlight differences rather than similarities. The ethnocentric approach, where the 'white' group is used as the ideal, and partial analyses are made of a limited number of disorders, has led to a misinterpretation of priorities. Ethnic minorities have similar patterns of disease and overall health as the ethnic majority. There are a few conditions for which minority groups have particular health needs, such as the haemoglobinopathies".

The 2004 Health survey for England built upon information obtained from the previous survey of 1999, with an increased number of households in order to boost the numbers of ethnic minority respondents, of the Black African population in particular. The data collected included interviews with a total of 12,644 adults over the age of sixteen, 5,828 from the general White British population and 6,816 from ethnic minority groups.

Self-reported data consistently shows some ethnic minority groups reporting poorer health than others, for example, in the 2004 Health Survey of England, 15% of Bangladeshi males described the state of their health as “bad or very bad” compared to 6% in the White British population. In contrast, males of Chinese and Black African origin report their state of health as better than that of the general population (The Information Centre, 2006).

1.9.1 Policies requiring or encouraging ethnicity data collection

The Race Relations Act was passed in 1976 criminalising racial discrimination in the work place and service delivery, and defines direct and indirect discrimination. The Race Relations (Amendment) Act 2000 placed key public bodies (including all government departments) under statutory duty to promote race equality which meant public authorities had to take steps to eliminate unlawful discrimination and promote equality of opportunity. The recent Equality Act 2010 harmonised and strengthened existing legislation to create a super act, making it criminal to discriminate or unfairly treat individuals on the grounds of nine protected characteristics, one of these being race (Equality and human rights commission, date of access 23/07/13. http://www.equalityhumanrights.com/legal-and-policy/equality-act/). Public authorities, such as the NHS, are required to make evident their compliance to the above legislation by publishing data and setting clear objectives for the future (Home Office, 2010).

1.10 Implications for this thesis and future research

Ethnicity data collection, analysis and use are complex but needed to determine if ethnic minorities are adequately represented in RCTs. The under-representation of ethnic minorities in clinical trials prevents the findings of clinical trials being generalised to the whole population and violates government polices promoting equality in opportunity and access to clinical trials.

The aims and objectives of the thesis stemmed from the challenges in ethnic minority data collection and analysis outlined above.

1.11 Aim and objectives

The aim of this thesis is to investigate ethnicity data collection in healthcare and the representation of ethnic minorities in RCTs, also to identify interventions to improve ethnic minority recruitment into RCTs and identify barriers to participation. This aim will be addressed through the following objectives:

Objectives

- Exploration of the UK ethnic minority population and how this has changed over 10 years
- Literature review of methods, interventions and barriers addressing the collection of ethnicity data in primary and secondary care
- Evaluation of the perceptions, experiences and willingness of lay South Asian volunteers to provide ethnicity data in healthcare situations

- Evaluation of health care professionals' perceptions and experiences of collecting ethnicity data in primary and secondary care
- Literature review of interventions to improve the recruitment and retention of ethnic minorities into RCTs
- Examining local datasets to assess completeness of ethnicity data collection and assess the representation of ethnic minority patients in RCTs utilising data from Hospital Episode Statistics and a Cancer Research Network
- Assess barriers to non-participation by ethnic group for cancer patients using Cancer Research Network data

The thesis consists of ten chapters. Chapter One has set out the background information and consequent challenges, building up to the aims and objectives. Chapter Two describes the methodology used throughout the thesis in order to address the specified aims and objectives. Chapter Three describes the changing composition of the ethnic minority population. Chapters Four, Five and Six focus upon ethnicity data collection and were conducted as part of the CRUK commissioned CanEth project. More specifically, Chapter Four reports the results of a systematic review of improving ethnicity data, Chapters Five and Six explore barriers to ethnicity data collection from a participant and healthcare professional perspective, respectively. This is followed by two chapters concerned with the participation of ethnic minorities in clinical trials. Chapter Seven explores the literature for interventions to improve ethnic minority participation in RCTS, whilst Chapter Eight reports the completeness of locally collected ethnicity then uses the data to ascertain whether under-representation is an issue in an ethnic minority rich Cancer Research Network (CRN). Chapter Nine describes my journey of establishing ethnicity data collection alongside trials recruitment data and explores

barriers to participation using the newly collected ethnicity data. To round off Chapter Ten discusses the research in its entirety, reports conclusions, limitations, further work and reflections.

Chapter 2 Summary of methods

The topics covered in this thesis were addressed from several different angles using a mixed methods approach. Quantitative, qualitative and epidemiological approaches were needed to tackle this undeveloped area of research, summarised below:

1. **Two systematic literature reviews**
    - Ethnicity data collection
    - Interventions to increase ethnic minority participation in RCTs
2. **Two surveys**
    - Healthcare professional's perceptions of ethnicity data collection
    - Reasons for choosing not to participate in a RCT by ethnic group
3. **Five focus groups**
    - With lay volunteers from the South Asian community
4. **Exploration/analysis of datasets:**
    - Census 2001 and 2011
    - Hospital Episode Statistics (HES) inpatient and outpatient
    - Cancer Research Network (CRN) trial recruitment data
    - CRN reasons for non-trial participation by ethnic group data

Further details of each method and their use are detailed as follows in this chapter.

2.1 Systematic literature reviews

I gained experience of systematic literature reviews in an earlier project exploring barriers and attitudes to participation in clinical trials (Ward et al, 2000). I learnt about literature searching and was able to replicate the original search strategy at a later date to capture more recent publications. Advice and guidance on systematic reviews, search strategy, terms, indexed subject headings and databases were provided by the Cochrane Methodology Review group, Warwick Centre for Evidence in Ethnicity Health and Diversity (CEEHD) specialist librarians and University of Warwick academic support librarian for biomedical sciences. Prior to commencing the reviews the following databases were checked to ensure this review had not already been conducted:

- Cochrane Database of Systematic Reviews (CDSR)
- Database of Abstracts of Reviews of Effects (DARE)
- TRIP database (clinical search engine)
- National Research register (this includes information about on-going reviews)
- Centre for Reviews and Dissemination (CRD) search filter with Medline (http://www.york.ac.uk/inst/crd/search.htm)

I attended systematic literature review training provided by the Cochrane Methodology Review group and the Centre for Review and Dissemination in York. These included methods by which the literature was identified, reviewed, extracted and quality checked.

The PICOS (Population, Interventions, Comparators, Outcomes and Study designs) framework was used to formulate the review questions by clarifying the scope of the review and devising the search strategy as follows:

- P=population or participants
- I=Intervention been trialled
- C=comparator between two or more alternatives
- O=Outcomes, quantitative or qualitative
- S=Study designs

Electronic databases were identified and search terms explored and developed using indexed subject headings and exploding key words. This process was performed for each database independently due to the variability of indexed subject headings per database. Keywords from relevant publications were also explored (Table 2.1). Adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guidelines ensured the reviews would be high quality and worthy of publication in peer reviewed journals.

Table 2.1 Search strategy development

1	The research question was broken down into component parts/concepts
2	A list of search terms for each concept was created through the use of synonyms, abbreviations, differences in spelling (transatlantic differences), identification of related terms
3	Exploding MeSH headings, trees helped develop the concepts and the list of search terms (indexed subject headings)
4	Use of truncation and wildcards
5	Relevant search terms for each concept were combined using operators: OR, AND, NOT
6	Limits applied included articles written in English, restricted articles about humans, and articles published between selected time points

2.2 Surveys

The survey method was used to gather the views of 1) health care professionals on ethnicity data collection and 2) ethnicity information and reasons given by cancer patient's choosing not to participate in a clinical trial. Both questionnaires were designed to be anonymous.

The healthcare professionals' survey was conducted as part of the CRUK Cancer and Ethnicity (CanEth) project and approved as part of the task by South Birmingham Ethics committee. The questionnaire was developed based upon a previous version by CEEHD at the University of Warwick and benefited from review by the CanEth project team. Perspectives of collecting ethnicity data were collected from both qualitative and quantitative angles.

A survey of cancer patients who declined participation in clinical trials was conducted in collaboration with a cancer research network. Questionnaires were developed jointly by me and the cancer research network team. Questions included basic information, such as gender, age group and disease site, as well as ethnicity data and an open ended question asking why the patient chose not to participate in the clinical trial.

2.3 Focus groups

The perceptions, experiences and willingness of South Asian volunteers to provide ethnicity data in primary and secondary care were explored through a series of

focus groups in the native language of each group and English where specified. Focus groups maximise interaction and debate between participants, and are a widely accepted method of gathering data to evaluate services, needs assessment and to conduct exploratory research in new areas (Schneider and Palmer, 2002). This research was conducted in collaboration with the Mary Seacole research centre at De Montfort University and the Ethnic Health Forum in Manchester as part of the CRUK CanEth project, using trained bilingual facilitators. A topic guide was developed as well as information sheets and consent forms and were granted ethical approval as part of the CanEth project. The facilitators used the topic guide which was specifically developed to focus on the five key areas of interest:

1. General opinions on the collection of ethnicity information
2. Experiences of providing ethnicity information
3. Categories used in practice
4. Language, religion and culture
5. How information should be collected

All sessions were recorded and transcribed by the trained facilitators.

Data generated by the focus groups were analysed by hand using a qualitative thematic approach. Data from the transcripts were examined, the accounts compared with one and another and common themes identified. Themes were developed and discussed by the project working group.

2.4 Datasets

I analysed four datasets as part of this thesis including census data, HES data from a large teaching hospital, clinical trials recruitment and reasons for non-participation data from a cancer research network.

Analysis of population level data was used to explore the structure of the ethnic minority population by age and gender for 2001 and 2011. I commissioned a dataset from Office for National Statistics (ONS) who collated information from the England and Wales, Scottish and Irish censuses of 2001. The dataset comprised of population counts by ethnic group, country of birth, age and gender. I used Microsoft Excel to analyse the data and calculated proportions of each ethnic group born within or outside of the UK. This was completed for each five year age band and gender with the resulting data used to construct population pyramids by ethnic group. The census data were used to assess the composition of ethnic minorities within the UK and West Midlands. The plots were then compared to existing population pyramids based upon 1991 census data to explore changes in the ethnic minority population (Owen 1996).

HES in and outpatient data were analysed using SAS statistical software (SAS Institute, Cary, NC) to assess quality and completion of ethnicity data across datasets. Further to this, I examined International Classification of Diseases version 10 (ICD-10) codes to identify disease sites and amalgamated smaller sites into the broader disease categories. Data from the inpatients and outpatients were merged

in SAS to form a final dataset which was used to calculate rates of cancer diagnoses by ethnic group.

Representation of ethnic minority patients in clinical trials was assessed using recruitment data from a cancer research network. Information from the analysis of the HES data described above was used to form a denominator for this purpose. Analyses were performed using Microsoft Excel with proportions of patients being recruited or not participating presented. Reasons for non-participation were further analysed using a qualitative thematic approach which I performed by hand.

This chapter provides a summary and justification of each method, more detail is provided in subsequent chapters. Chapter Three presents data from the analyses of the 2001 Census.

Chapter 3 The changing structure of the UK ethnic minority population

The aim of this chapter is to provide an overview of the ethnic minority population in the UK and illustrate how the age structure of the ethnic minority groups has changed over 10 years by comparing it to data from 1991 census. Only basic information has been presented from the 2011 census as only limited data were available at the time of writing this thesis.

3.1 Census 2001

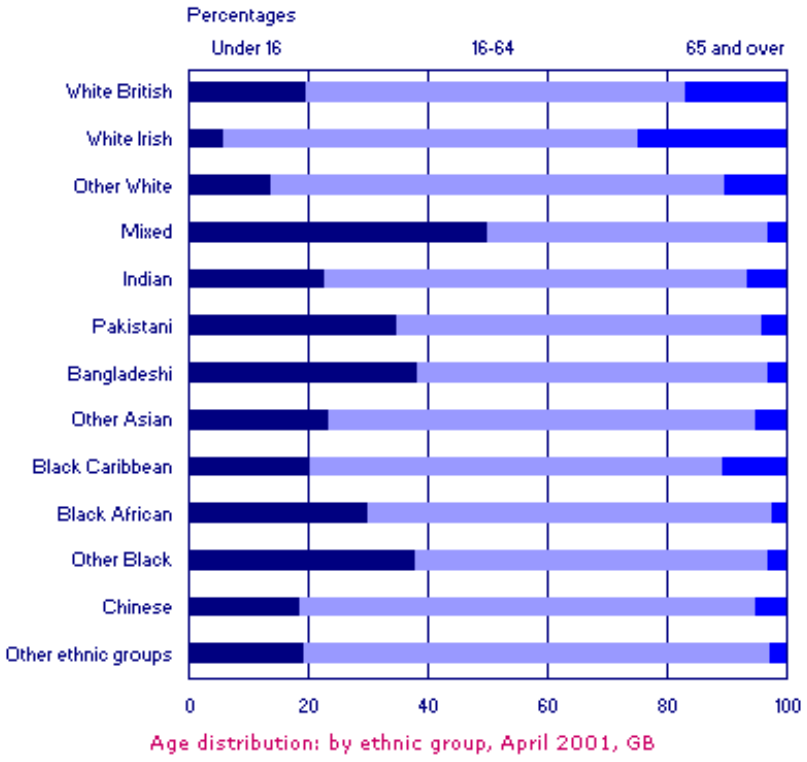
Ethnicity data available in the 2001 census were explored by age, gender, country of birth and nationality, language, religion and geographic distribution of ethnic minorities.

3.1.1 Age structure

As expected the non-White population of the UK was younger than the White population. This can be explained by the timing in which these groups migrated to the UK and higher birth rates in these groups (Figure 3.1). The 'mixed' group was considerably younger than all the other groups with approximately 50% being under 16 years of age. This was followed by 38% under 16 years in the Bangladeshi and Other Black groups compared to 20% in the White group (Office for National

Statistics, 2006). The Chinese are the only non-White population with an 'under 16 years' group smaller than the White British group. All ethnic minority groups have a distinctly smaller older population (65+ years) than the White population which ties in with the passing of the British Nationality Act in 1948, 53 years previous. An increase in the older non-White populations is expected in the future, however we may need to adjust for 'salmon-bias' (which refers to people returning to their country of origin post retirement, Bhopal, 2007).

Figure 3.1 Age distribution by ethnic group, 2001/2002



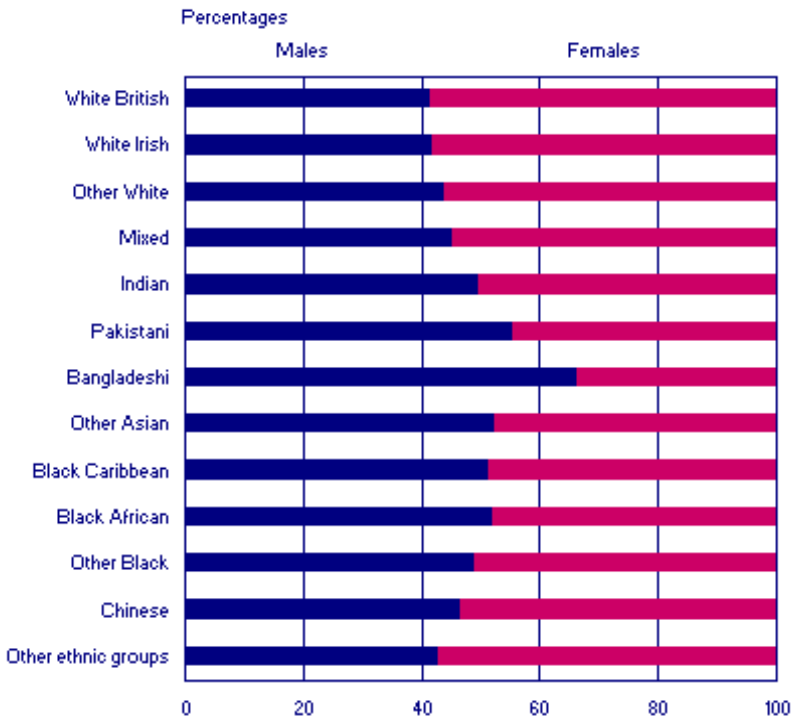
(Source: Census, April 2001, Office for National Statistics; Census, April 2001, General Register Office for Scotland)

3.1.2 Gender

In the White population women outnumber men in the older age brackets (65 years plus) due to differences in life expectancy and mortality rates. In 2001, this was 58% vs. 42% respectively (Figure 3.2). The same pattern was also apparent in Mixed and Chinese groups but was reversed in the Bangladeshi, Pakistanis and Other Asian groups with 34%, 45% and 48% females respectively (Office for National Statistics, 2006). The assumption that there are more South Asian male immigrants than females could explain this anomaly.

(Source: Census, April 2001, Office for National Statistics; Census, April 2001, General Register Office for Scotland)

Figure 3.2 Sex distribution of people 65 & over: by ethnic group, April 2001, GB



3.1.3 Country of birth and nationality

Country of birth and nationality differ from one another for two important reasons. Firstly, people over the age of 18 years, who were born abroad and have been living in the UK for a minimum of 5 years, are eligible to apply for British Citizenship (UK Border Agency). This period was reduced to three years where an individual is married to a British citizen. Secondly, people born abroad can have British nationality; this is often the case of children born to parents serving in the military. In 2007, 266,000 people reported Germany as their country of birth, even though 70% of these were UK nationals (Ellis, 2009).

Reports based on the Annual Population Survey show the non-UK born population between 2004 and 2007 to have increased by estimated 1.1 million. The largest group in both 2004 and 2007 accounting for approximately 10% were born in India. The population born in Poland rose sharply from 95,000 in 2004 to 405,000 in 2007. Poland joined the European Union along with eight other Central and Eastern European countries in May 2004, which could explain this increase, see Table 3.1.

Table 3.1 Five most common non-UK countries of birth in the UK, 2004 and 2007 (thousands)

	2004		2007	
	Country	Estimate	Country	Estimate
1	India	502	India	613
2	Republic of Ireland	452	Republic of Ireland	420
3	Pakistan	281	Poland	405
4	Germany	275	Pakistan	377
5	Bangladesh	225	Germany	266

(Source: Office for National Statistics, Populations trends, Spring 2009)

3.1.4 Language

Language was not available in the 2001 census, however, results from the 'Health Survey of England-the health of minority ethnic groups' showed only 20% of Bangladeshis reporting English as their main language, the majority of these being British born or young migrants (Department of Health, 1999). Groups most likely to speak English were young males or people who had attained some level of education, over 50% of Bangladeshi men and women reported not holding any qualifications. A survey reported large proportions of non-White ethnic groups to be below what was referred to as a "survival level of competence", defined as the ability to read a phone book or school report (Carr-Hill et al, 1996). Rates were lowest amongst the Bengali group, with only 14% of those not born in the UK attaining this level. This level improved in the Punjabi (26%) and Gujarati groups (29%). The Chinese group were reported to have the highest rate, with 49% reaching the "survival level of competence".

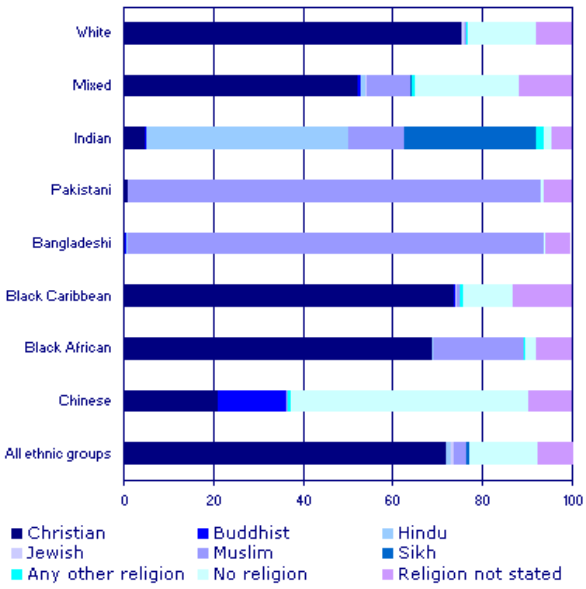
3.1.5 Religion

Religion was included for the first time as an optional question in the 2001 UK Census. Almost 8% of the population did not respond to this question (Office for National Statistics, 2005). Of those who responded, the majority (72%) of the White British population classified themselves as Christians; proportions were similar for the Black Caribbean group (74%) and Black African groups (69%) but lower in the Mixed group (52%, see Figure 3.3). Approximately, 15% of the population classified

themselves as having 'no religion', of these 53% were of Chinese ethnicity and 23% of Mixed ethnicity. Only a small proportion of Pakistanis and Bangladeshis classified themselves as having 'no religion' (0.56% and 0.45% respectively).

Muslims made up the largest non-Christian group with 52%, followed by Hindus (18%), Sikhs (11%), Jews (9%) and Buddhists (5%). A large proportion of the Muslim group comprised of people originating from South Asian countries, such as Pakistan and Bangladesh (where 90% of the population identify as Muslim). In addition, 20% of Black Africans and Other ethnic groups classified themselves as Muslim.

Figure 3.3 Religious composition of ethnic groups in England, Scotland and Wales, April 2001



Sources: Census, April 2001, Office for National Statistics; Census, April 2001, General Register Office for Scotland

3.1.6 Geographic distribution of the ethnic minority population in the UK

In 2001, 96% of the non-White ethnic population were reported to reside in the busy urban areas of England, pre-dominantly London, West Midlands, West Yorkshire and Greater Manchester. London as the capital was appealing to many immigrants and was home to 45% of the non-White population, 11% resided in the West Midlands and 5% in West Yorkshire and Greater Manchester. Migrants of Pakistani origin were well dispersed between the four regions, with 19% in London, 21% West Midlands, 20% Yorkshire, 16% Greater Manchester. In contrast, the Bangladeshi and Indian populations were concentrated in London (54% and 40% respectively) (CRE, 2007, Office for National Statistics, 2001). Furthermore, London was inhabited by large proportions of other ethnic minority groups, for example, 61% of the Black Caribbean population and 78% of Black Africans.

3.2 Commissioned census 2001 dataset

A dataset containing ethnic group and country of birth by age and sex was commissioned for the UK through the ONS. Information from the England and Wales, Scotland and Northern Ireland 2001 censuses were collated by ONS. Ethnicity codes varied in each country but there was considerable overlap (Table 3.2). Ethnic group codes were harmonised to create 11 'UK' ethnicity groups which were used in the analyses (Table 3.3).

Age data were split into five year bands from 0 to 84, with ages 85 and over combined into a single category. Country of birth was split into a binary category (UK born and non-UK born). Data were available for the UK and broken down to local authority level for England and Wales.

Table 3.2 Ethnic group categories for 2001 census by country

England And Wales	Scotland	Northern Ireland
01 British	01 White Scottish	01 White
02 Irish	02 Other White British	02 Irish Traveller
03 Other White	03 White Irish	03 Mixed
04 White and Black Caribbean	04 Other White	04 Indian
05 White and Black African	05 Any Mixed Background	05 Pakistani
06 White and Asian	06 Indian	06 Bangladeshi
07 Other Mixed	07 Pakistani	07 Black Caribbean
08 Indian	08 Bangladeshi	08 Black African
09 Pakistani	09 Other South Asian	09 Other Black
10 Bangladeshi	13 Chinese	10 Chinese
11 Other Asian	10 Caribbean	11 Other Ethnic Group
12 Black Caribbean	11 African	12 Other Asian
13 Black African	12 Black Scottish or Other Black	
14 Other Black	14 Other Ethnic Group	
15 Chinese		
16 Other Ethnic Group		

Table 3.3 Harmonising ethnicity categories to create UK ethnic groups code

UK ETHNIC GROUPS	EW Codes used	NI Codes used	S Codes used
White	01; 02; 03	01; 02; 03; 04	01; 02
Mixed	04; 05; 06; 07	05	03
Indian	08	06	04
Pakistani	09	07	05
Bangladeshi	10	08	06
Other Asian	11	09	12
Black Caribbean	12	10	07
Black African	13	11	08
Black Other	14	12	09
Chinese	15	13	10
Other	16	14	11

(Office for National Statistics, 2013)

Once the ethnic codes were collated and data tables constructed by ONS, I was able to use these data to explore the population structure of UK ethnic minorities.

3.2.1 Population pyramids

Population pyramids of the age and gender structure of the UK were constructed for grouped populations, such as all the UK, Whites and non-White minorities as well as for ethnic minority groups independently. Proportions of male and female populations were calculated according to country of birth and displayed in five year age bands with the youngest at the base and the oldest at the top. Births outside of the UK are presented as dark shading.

The population structure of the whole population and the White group were very similar (Figures 3.4a and b). The shape displays slow population growth with age bands of roughly equal size with a rapid shrinkage post retirement age. It was also possible to see the imbalance of males and females due to longer female life expectancy as well as the post-war baby booms. In contrast, the population structure of the ethnic minorities starts with a broad base and gradually diminishes to a small apex creating a pyramid like shape (Figure 3.4c). This is reflective of the large numbers of children and young adults and sparse older population.

Age and sex pyramids with shading to indicate the proportions of the population not born in the UK, plainly demonstrates the younger age structure of the ethnic minorities compared to the White populations of 1991 and 2001 (Figures 3.4b and c). Figure 3.4c clearly show two distinct sections revealing a non-UK born shaded top half, in contrast to a bottom half made up of their UK born offspring. Over 50% of ethnic minorities in the UK in 2001 were born in the UK, which means that variables such as country of birth which were once informative were no longer effective indicators of ethnicity.

In the Pakistani and Bangladeshi groups an age imbalance can be observed with a much smaller proportion of females to males (Figures 3.4g and h). This may be because during migration it was the males that relocated first in order to secure employment and housing, families followed later. This gender imbalance was also present in the Black Caribbean population in 1991 where men outnumbered the women but was not evident from the 2001 data (Figure 3.4d). The structure of the

Black ethnic groups forms a pyramid like shape with the addition of two protrusions or bulges. The first represents the influx of early migrants and the second larger protrusion represents the British born children of these migrants. In the African group it is possible to see the base broadening out once more as a result of a growing third generation.

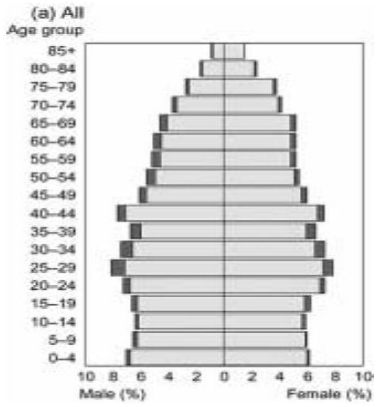
The South Asian group populations are very pyramid-like having a wide base narrowing to a small apex (Figures 3.4f, g and h). There are more males than females in the older age groups. Both the Pakistani and Bangladeshi populations are widest at the base indicative of a large number of children and young people, unlike the Indian population where the groups are roughly equal into the middle age groups. The Chinese population structure tells a different story. The base is much narrower than that of the other groups (Figure 3.4i). The largest group was the young adults, most likely overseas university students.

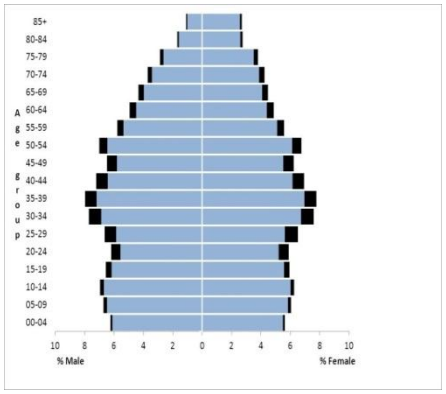
Comparison of the 1991 and 2001 population pyramids particularly the two bands (0-4 and 5-9 years), reveals a decrease in proportions of younger populations compared to previous years. Decreasing band widths were displayed in the charts of the whole population, Whites, Black Caribbean's, Indians and Chinese groups (Figures 3.4a, b, d, f and i). Bands were seen to widen in the ethnic minorities as a whole, in particular in the Black Africans, Pakistanis and Bangladeshi groups (Figures 3.4c, e, g and h).

Figure 3.4 Age and sex distribution of persons born within and outside of the UK by ethnic group 1991 (Owen, 1996) compared to 2001

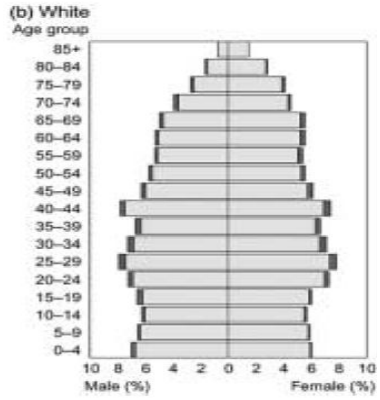
Note: persons darker shading represents born outside the UK

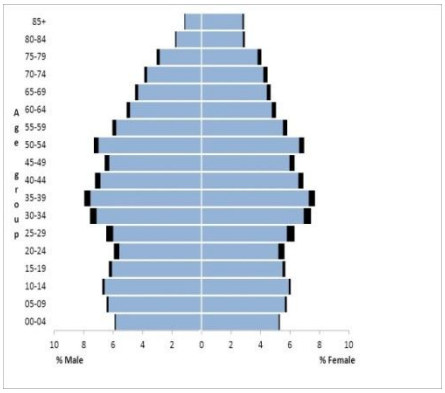
a) All 1991 2001





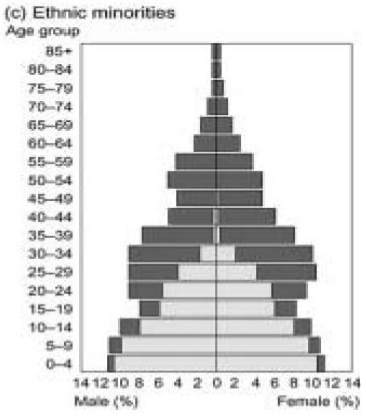
b) White 1991 2001

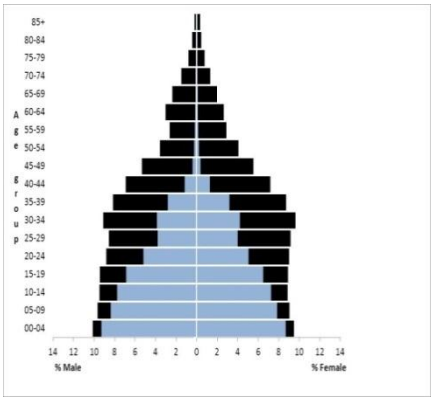




c) Non-white ethnic minorities

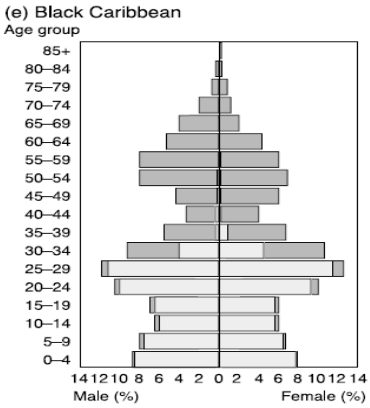
1991 2001



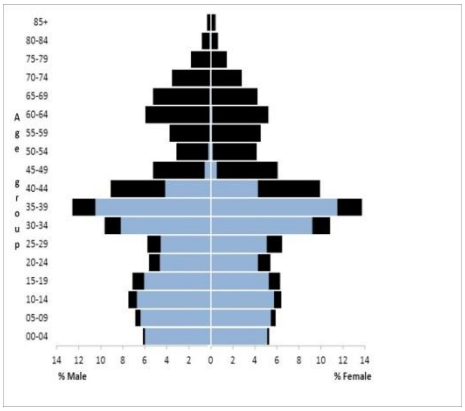


d) Black Caribbean

1991

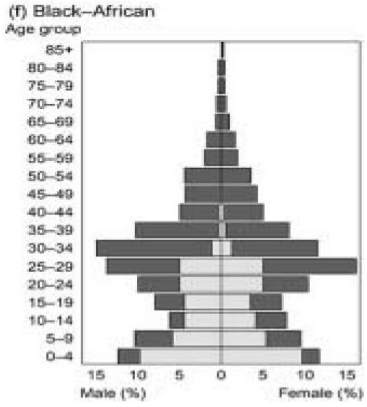


2001

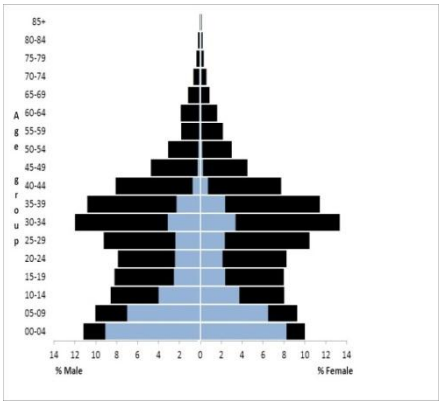


e) Black African

1991

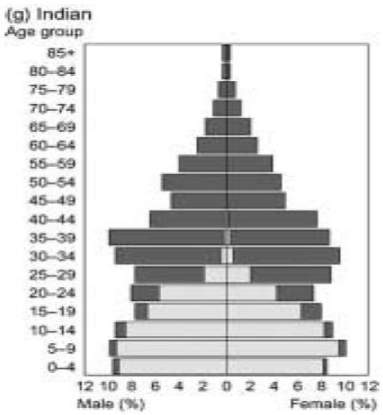


2001

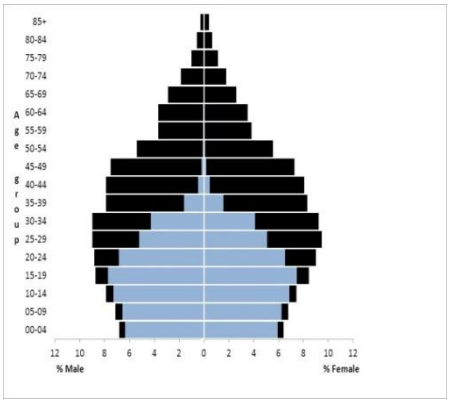


f) Indian

1991

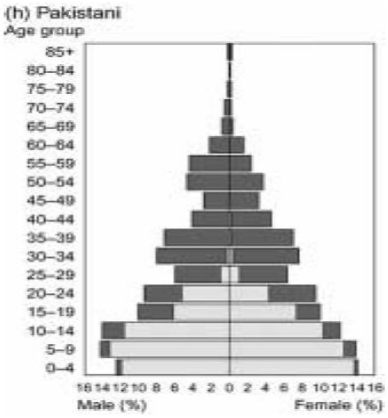


2001

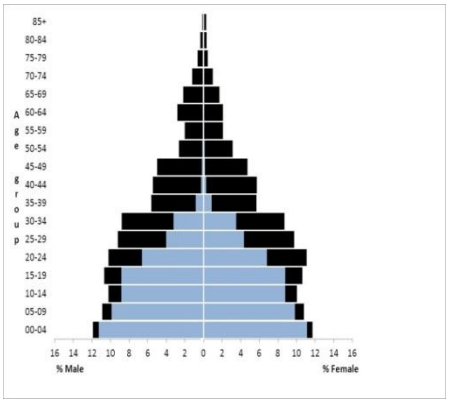


g) Pakistani

1991

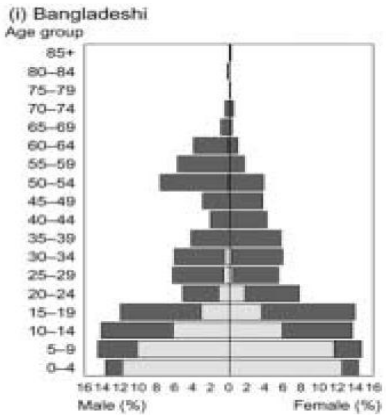


2001

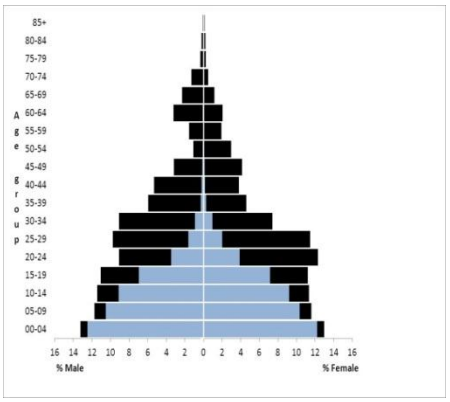


h) Bangladeshi

1991

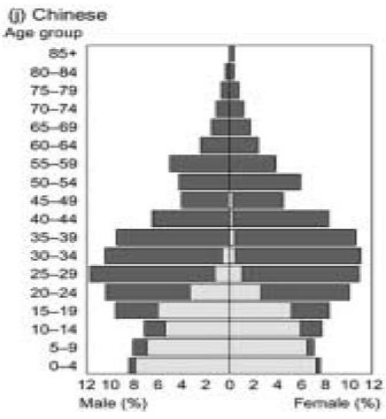


2001

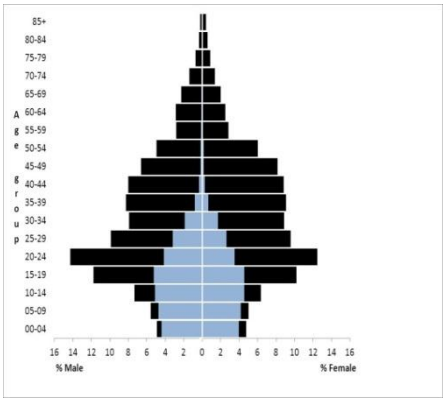


i) Chinese

1991



2001



Ethnic minorities are getting older and increasing in number. Disease burden will rise and health services will need to adapt to meet the needs of the ageing ethnic minority population. However, it will be difficult to plan given the current poor collection of ethnicity data. In the next chapter, I present results from a review of the evidence base to see how we can improve ethnicity data collection in healthcare and present examples of good practice.

Chapter 4 Systematic literature review 1: Methods, interventions and barriers addressing the collection of ethnicity data in healthcare

Ethnicity data is known to be poorly collected in healthcare. Although there have been significant improvements over recent years the data remains incomplete and of poor quality (HES online, 2004 and HES online, 2011). This literature review aims to identify methods, interventions and barriers to the collection of ethnicity data (or ethnicity profiling) in primary and secondary care. The review was conducted as part of CanEth, the CRUK commissioned feasibility project.

4.1 Methods

Literature searches were carried out in Embase, Psychlit, MEDLINE, PsycINFO and CINAHL databases in conjunction with the Warwick Centre for Evidence in Ethnicity Health and Diversity (CEEHD) with the aid of an information scientist specialising in ethnicity. The searches aimed to capture literature pre- and post the passing of the Race Relations Act (Amendment) in the UK in 2000, and the USA National Institute for Health Revitalisation Act of 1993 which prompted interested in the inclusion of ethnic minorities in clinical trials (NIH, 2001). The searches were also limited to

articles and reports written in the English language due to the time limitations of this thesis and the CanEth project.

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 strategy

The search terms were developed in conjunction with the Centre for Evidence in Ethnicity Health and Diversity (CEEHD) ethnicity information specialists at Warwick, who also advised on the choice of databases. The search strategy focused on three key concepts:

1) Ethnicity
2) Data collection or data monitoring
3) Disease sites (cancer or other chronic or long term diseases such as stroke, diabetes, coronary heart disease)

Table 4.1 presents free text and Medical Subject Headings (MeSH) indexing terms identified for each concept.

Table 4.1 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]		

(Iqbal et al, 2009)

The first literature search was limited to 2000-2007 with the aim of identifying recent literature and good practice guidelines. The literature search was re-run using the same search terms but limited to literature published between 1990-1999 to ensure all relevant articles around the time of the 1993 Revitalization Act were captured (NIH, 2001). A third cancer site specific literature search was conducted with no restrictions on the year of publication, which included cancer sites known to have differing risks by ethnic group (e.g. Breast, Prostate, Head and Neck, Colorectal and Cervical malignancies). Search terms for ethnicity and data collection were used as previously described whilst terms related to the cancer site of interest were added. Cancer site specific searches were conducted in each database for each disease site independently. These were then concatenated. Duplicates were removed using the Endnote software function and further by hand.

4.1.2 Selection criteria

Selection and inclusion and exclusion criteria for the populations of interest, interventions and study designs were developed prior to commencing the review (Table 4.2). I was already aware from a scoping search that many articles used ethnicity but did not report how the information was collected. Articles such as this and those discussing the need to collect ethnicity data were excluded.

Table 4.2 Selection criteria and inclusion and exclusion criteria

Selection criteria	Inclusion criteria	Exclusion criteria
Population	Patients from a Black, Minority or Ethnic group. This includes indigenous populations of invaded countries, e.g. Australia and USA	Black, minority and ethnic groups in their countries of origin
Interventions	Any method, intervention or barriers addressing ethnicity data collection	Methods, intervention or barriers NOT addressing ethnicity data collection OR use of ethnicity but no explanation as to how it was collected
Study design	Any	Opinion pieces discussing the importance of collecting ethnicity data and reporting collecting data

4.1.3 The selection process

The review process was broken down into three stages: 1) title, 2) abstract and 3) article review. Categories emerged during the title screening process and were used to categorise abstracts and articles (Table 4.3).

Table 4.3 Coding used during screening process

Code	Description
M	Definitely/possibly contains methodology of ethnicity data collection
U	Use of ethnicity data but no explanation of how the data were collected
I	Insubstantial
D	Duplicate
N	Not relevant

The following steps were taken in order to minimise selection bias:

- Two reviewers independently screened titles, abstracts and articles
- Two reviewers independently applied the inclusion criteria to the full articles. Any differences were resolved by consensus or referral to a third reviewer
- Two reviewers independently extracted data from the final selection. As before, any differences were resolved by consensus or referral to a third reviewer

4.1.4 Grey literature review

Grey literature was searched as not all research is published, or all published research indexed in the main databases, nor do all databases index consistently. In addition not all authors describe their research using the same terms.

Publication types considered

- Conference proceedings/abstracts
- Theses
- Reports, government documents
- Key websites, such as the Kings Fund and Commission for Racial Equality

Grey literature searches were conducted using the keywords in the search, such as: ("data collection" OR "data monitoring") AND ("ethnic" OR "ethnicity"). The searches were conducted in Google and Google Scholar. Due to the large volume of results obtained using this method, screening was restricted to the first 50 pages only. The majority of grey literature was rejected because 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 methods, policies or procedures, or those providing guidance to improve ethnicity data collection 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
- (http://www.lho.org.uk/)
- National Cancer Library
- (http://www.library.nhs.uk/cancer/)
- Office for National Statistics
- (http://www.statistics.gov.uk/)
- Department of Health
- (http://www.dh.gov.uk/en/index.htm)

4.1.5 Data extraction

The data extraction form was designed to capture the following information:

1. Type of cancer
2. Country of study
3. Ethnic group
4. Type of study
5. Focus of the study
6. Key findings

4.2 Results

Due to the nature of the evidence, published literature was reviewed using different steps to grey literature. The published papers were screened in three stages – 1) title review, 2) abstract review and 3) article review, and the grey literature reviewed by full article. The number of hits from the grey literature searches were too many and too duplicated to be able to present rejected articles in a meaningful way (e.g. links to the same report listed on multiple websites), therefore only the relevant articles are presented.

4.2.1 Published Literature

The systematic review of the published literature produced a total of 2404 'hits'; 720 for the period 1990-1999 and a further 1684 for 2000-2007 (Table 4.4).

Table 4.4 Search 'hits' by database*

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

* Numbers excluding duplicates are shown (total considered 2658). (Iqbal et al, 2008)

Upon review of the 2404 selected titles, only 322 appeared to suggest that they discussed the methodology of either collecting or monitoring ethnicity data. A full review of these 322 abstracts revealed only 26 potentially fulfilling the inclusion criteria, Table 4.5. The majority of articles rejected (57%) were due to the article using ethnicity data as opposed to the methods for collection.

Table 4.5 Stages 1 and 2: Title and abstract review

			Abstract category*					
Period	**No of titles**	**No of Abstracts**	**M**	**U**	**I**	**D**	**N**	**No meeting criteria**
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 26 potential articles which assessed ethnicity data collection or ethnicity monitoring were reviewed in full; only 19 of these included information about data collection or monitoring. Once again, the main reason for rejecting articles was due to the 'use' of ethnicity and not data collection or monitoring (Table 4.6). Please note, one of the potentially relevant papers was included based upon the abstract only as the full paper was unavailable at the time of review.

Table 4.6 Stage 3 article review

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

*M=Methodology, U=Use, I=Insubstantial, A=Abstract only (Iqbal et al, 2008)

Owing to the low numbers of papers fulfilling the inclusion criteria for acceptance and the interest in cancer sites given the research was funded by CRUK, the search was repeated for specific cancer sites as shown in Table 4.7.

Table 4.7 '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

(Iqbal et al, 2008)

This resulted in 539 potential articles of which 469 were deemed to possibly fulfil the criteria specified for acceptance following the screening of titles and abstracts as shown in Table 4.8 (Iqbal et al, 2008).

Table 4.8 Cancer site specific Stages 1 and 2 review

			Abstract category*					
Disease site	**No of titles**	**No of abstracts**	**M**	**U**	**I**	**D**	**N**	**No of articles**
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 four articles of the 469 abstracts reviewed discussed ethnicity data collection or monitoring, with two of these having been identified in the previous non-disease site specific searches. The remaining two papers did not involve data collection or monitoring upon full review of the article (Table 4.9) (Iqbal et al, 2008).

Table 4.9 Cancer site specific Stage 3 article review

		Article category*				
Disease site	**No of articles**	**M**	**U**	**I**	**A**	**No of articles remaining**
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

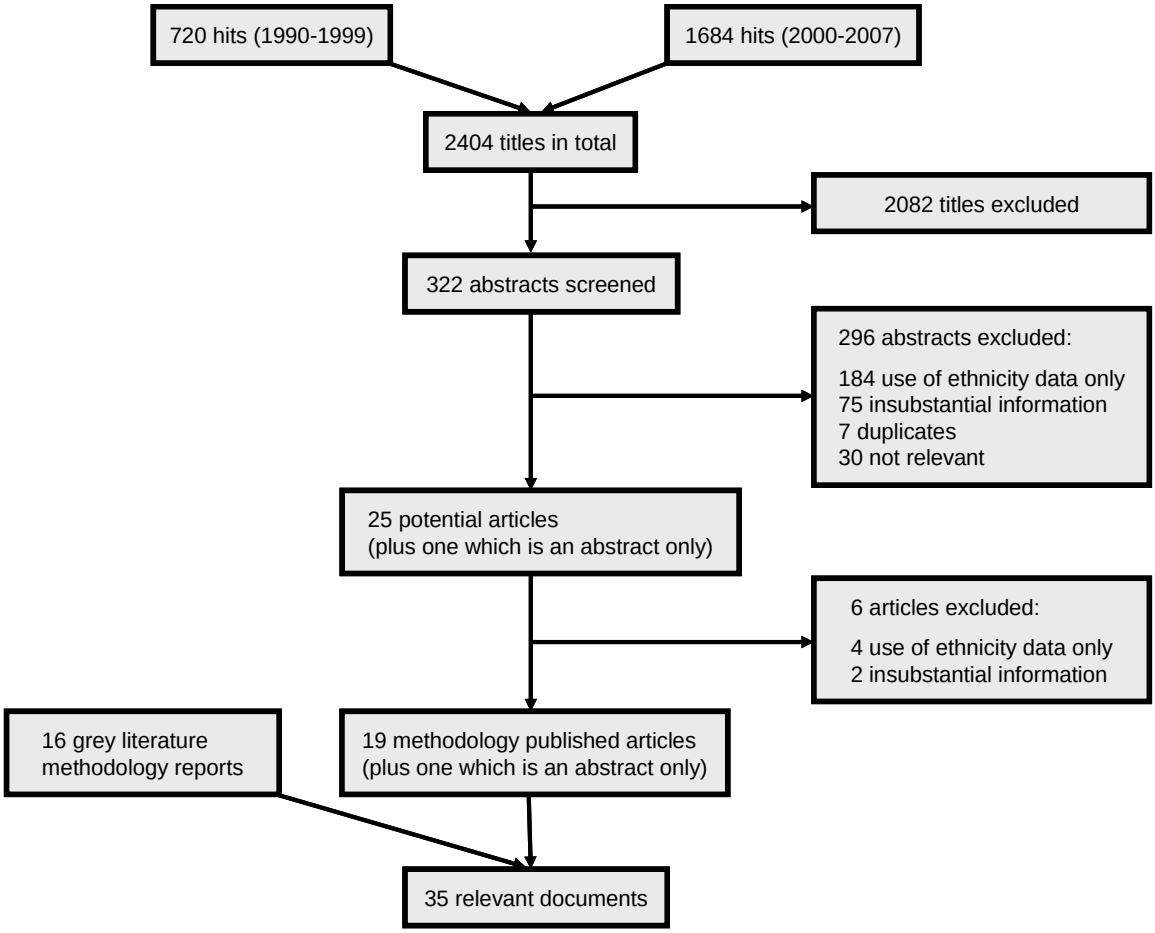
4.2.2 Grey literature

Searches on key websites and Google identified a large amount of information linked to ethnicity. Upon review, 53 reports were identified as being possibly linked with the collection or monitoring of ethnicity data. The main reason for rejection of potentially relevant 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 reports screened, 16 were selected for inclusion in the review (Figure 4.1).

4.2.3 Summary of literature searches and reasons for rejection

The paucity of evidence identified by the literature review highlights the emerging nature of research into this area. Of the 2404 possible published articles, only 19 (0.8%) were selected for inclusion in this review (Figure 4.1). The main reason for rejection was articles using ethnicity (57%) but providing no description of the methods used for ethnicity data collection. Ethnicity 'use' included research aimed at specific ethnic groups, e.g. 'Chinese' or 'South Asian' for health promotion or interventions.

Figure 4.1 Ethnicity data collection and monitoring review selection process



(Iqbal et al, 2009)

Many articles used the term 'multi-ethnic' to describe their study populations where ethnicity had not been recorded but participants were considered to be a representative sample, e.g. smoking cessation in a multi-ethnic deprived area. Articles such as this were rejected where they did not provide any information about ethnicity data collection. Other reasons for rejection included published

articles which discussed the population of their own country, such as multiple births in Saudi Arabian desert climate.

A systematic review by Ma et al (2007) similarly found a lack of reporting of the methods of ethnicity data collection when examining reporting practice of race/ethnicity and socioeconomic status in biomedical journals. Additionally, there was a large amount of grey literature but upon review only 16 reports were considered relevant to this review. Of the articles included in the review, 19 (54%) were identified from published literature with an additional 16 (46%) from grey literature. Data extracted from the selected published and grey literature are presented in tables 4.10 and 4.11 respectively.

Table 4.10 Summary of published articles

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 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.
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)
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. Train the trainer 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

Author , year Of publication	Type of cancer	Country of study	Ethnic group	Type of study	Description of content	
					Focus of the study	Key findings
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
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
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	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 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

Author , year Of publication	Type of cancer	Country of study	Ethnic group	Type of study	Description of content	
					Focus of the study	Key findings
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
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.
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
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

Author , year Of publication	Type of cancer	Country of study	Ethnic group	Type of study	Description of content	
					Focus of the study	Key findings
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
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
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
Frost, 1994	Non-cancer	USA	American Indians Alaska Natives	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 died in Washington are more likely to be coded correctly

Author , year Of publication	Type of cancer	Country of study	Ethnic group	Type of study	Description of content	
					Focus of the study	Key findings
Sugarman, 1993	Non-cancer, End stage renal disease	USA	American Indians Alaska Natives	Audit	Misclassification of American Indians ad Alaska natives on the Renal Disease Stage register and the impact upon disease statistics	Ethnicity validated against the Indian Health Service database 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

(Iqbal, et al, 2008 and 2009)

Table 4.11 Summary of grey literature

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 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 the race, ethnicity, and primary language data to improve quality of care 7.How to train staff to collect this information
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.Datanet system aids use of collected data

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	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
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 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 ethnic data in secondary care services

Title, year of publication	Authors	Country of report	Ethnic groups	Type of research	Description of content	
					Focus(es) of report	Key findings
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 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
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: 1.Standardise method of collection (self-report should be used whenever possible) 2.Point of data collection (admission) 3.Standardise 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

Title, year of publication	Authors	Country of report	Ethnic groups	Type of research	Description of content	
					Focus(es) of report	Key findings
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, 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. Advises against transferring ethnicity from one form to another.
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 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, 12% did not give 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	2 methods are proposed, one question (ethnicity) and 2 question (ethnicity and nationality). 2 question method should be used whenever possible

Title, year of publication	Authors	Country of report	Ethnic groups	Type of research	Description of content	
					Focus(es) of report	Key findings
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
Ethnic monitoring: A guide for public authorities, 2002	Commission for Racial Equality	UK	All	Guidelines	Ethnicity 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.
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.

Title, year of publication	Authors	Country of report	Ethnic groups	Type of research	Description of content	
					Focus(es) of report	Key findings
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
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 4 sections: personal details, patient satisfaction, health and ill health, ethnic classification. The data has been used to inform planning strategies, detailed in the report.
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

(Iqbal, et al, 2008 and 2009)

4.3 Synthesis of results

Seven themes emerged from the systematic literature review as listed:

1. Ethnicity data collection and monitoring
2. Categories for defining ethnic group
3. Other Indicators of ethnicity
4. Methods of collection
5. Barriers to collection
6. Interventions
7. Data quality and completeness

Each identified theme is discussed in more detail below.

4.3.1 Ethnicity data collection and monitoring

The following six guidelines and reports providing examples of good practice were the most useful in determining methods for ethnicity data collection and monitoring:

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

These grey literature reports presented the best practice evidence for ethnicity data collection and coding, addressing barriers to ethnicity data collection, interventions to combat incompleteness of data, best practice policy, the need for complete ethnicity data and training of healthcare professionals to collect these data. However, none of the above reports discussed the validation of ethnicity data or the use of name recognition programs as a validation tool. This may be due to the consensus that self-reported ethnicity is the gold standard. Several of the above reports discussed the collection of self-reported ethnicity data at a single time point, which was then verified at subsequent visits. The Department of Health 2005 guidelines presented examples of best practice across the UK (Department of Health, 2005). A common barrier to ethnicity data collection which affects both patients and healthcare professionals was the concept that ethnicity data was collected but not utilised. Key reports where ethnicity data collection/monitoring has been successfully conducted stated, demonstrating how the collected data will be used, is just as important to improved collection as adequate resources and awareness and training (Race for Health, 2006; Regenstein and Sickler, 2006).

In a report which aimed to investigate the state of ethnicity data collection in US hospitals through the use of surveys and hospital site visits revealed that a lack of standardisation made this difficult in practice, although there were good levels of commitment to ethnicity data collection (Hasnain-Wynia et al, 2004).

Weinick et al (2007) described the lessons learnt from executing a publicly mandated racial and ethnic data collection program in Boston and Massachusetts

acute care hospitals. The authors recommended data collection tools to be standardised across hospitals whilst still catering for the needs of individual hospitals and allowing patients a choice in how they choose to describe themselves. These findings were confirmed by Ford and Kelly (2005) who also stressed the importance of consistent methods of collection and the need for improved data collection tools.

'Who, when, and how: The current state of race, ethnicity, and primary language data collection in hospitals' report (Hasnain-Wynia et al, 2004) made five recommendations for improving ethnic data collection and quality as listed below:

- Hospitals need to standardise the method of collection (self-identification should be used whenever possible)
- Point of data collection, e.g. on admission/registration recommended
- Categories used for race and ethnicity should be the same across the board, ideally census but refinement is supported as long as data can be combined into census categories
- 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
- Patient concerns should be addressed prior to collection, response to concerns and explanations of data use should also be standardised

The impact of accurate ethnicity data collection has not yet been fully recognised and there is still a long way to go before the data are complete and reliable (Hasnain-Wynia et al, 2004, 2001a). It is important to collect accurate ethnicity data for planning and policy making reasons.

4.3.2 Categories for defining ethnic group

Ethnic groups may vary greatly between countries and therefore categories differ depending on the country where the research is carried out. A report looking at ethnicity questions asked in census surveys in and across the United Nations found a total of 107 questions were being asked by 95 countries collecting ethnic group as part of their census. Only 12% of the countries collecting ethnicity data offered categories for 'mixed' identities' or allowed for multiple boxes to be selected (National statistics, 2003). Other international guidelines specified that the gold standard categories could be expanded to cater for local populations as long as they could be concatenated back to the standard national categories (Race for health, 2006; Weinick et al, 2007; Commission for Racial Equality, 2002).

Inconsistencies were found in response type being used in connection with ethnicity questions in the censuses, five different types of response category were identified by the United Nations report as follows:

- Coded categories with text boxes
- Coded categories without text boxes
- Yes/No questions

- Free text self-report
- Unknown response

The UK gold standard categories were taken from the census 'ethnicity' question. Ethnicity was first collected in the 1991 census when 10 categories were collected, this was later expanded in 2001 to the 16+1 categories which saw the introduction of a 'mixed' category and expansion of the 'White' category (as described in Chapter One). The Commission for Racial Equality (CRE) report (2002) and the Department of Health (DoH) guide to ethnic monitoring (2005) both stressed the importance of not allowing patients to see the 'not stated' category on paper forms or offering the patient this option verbally, nor misled the patient into thinking the question is compulsory. Another key difference between the 1991 and 2001 census ethnicity categories was the ability to state ethnicity in more detail by writing in free text field where 'other' was selected within any ethnic group category.

4.3.3 Other indicators of ethnicity

The UK DoH guidelines (2005) encouraged the collection of religion through the use of seven categories (Christian, Buddhist, Hindu, Jewish, Muslim, Sikh, any other religion, not stated). The ONS went one step further and distinguished between religious practice (active faith and participation in religious activity and worship) and religious identity (belonging to a religious community even if the religion is not necessarily practiced). Special attention was given to the main religious groups: Christian, Buddhist, Hindu, Jewish, Muslim and Sikh as people from many diverse backgrounds make-up these religious groups.

The collection of information about diet as five categories was actively encouraged (no special requirements, vegetarian, vegan, restriction according to religion, e.g. Kosher or Halal, and food allergies or medical conditions). The term 'Kosher' is used to describe food prepared according to Jewish ritual; 'Halal' is used to describe meat prepared in accordance with Muslim law. Both terms 'Halal' and 'Kosher' are listed in the Oxford English dictionary.

Language was believed to be important to record so interpreters can be called for if needed. Relatives were not deemed to be adequate interpreters as they may not have the vocabulary or understand the technical nature of questions, especially when in a medical setting. There is also the uncertainty of not knowing whether you have been told everything or if the information has been edited, this may certainly apply when intensely personal questions are asked. There are also problems when patients need to make decisions because it is difficult to know who made the decision.

The ONS recommended Nationality also be collected, the six suggested categories were: English, Scottish, Irish, Welsh, British and Other (National Statistics, 2003). Responses should be re-ordered according to where the question is asked, e.g. English should be at the top of the list when asked in England.

The Individual Patient Registration Profile (IPRP) developed by Lambeth Primary Care Trust collected self-reported ethnicity, religion, language and need for an interpreter in addition to the routinely collected data items, such as gender, date of

birth and smoking status. Ethnicity group codes were expanded to meet the needs of the local population, but this was done in a way that allowed the categories to be concatenated down to the 16 census 2001 categories when required (Race for Health, 2006). The IPRP is an on-going project which attempts to collect data on a central database (Datanet) and includes a system which allows the collected ethnicity and profiling data to be used in research projects. Central Liverpool NHS Primary Care Trust also conducted a similar patient profiling exercise and collected detailed ethnicity data including 'spoken' and 'written' language, as well as additional information about general health (Adebayo and Mitchell, 2004).

Indicators of ethnicity and their usefulness were discussed by Gill et al (2006) as part of the needs assessment series. Country of birth has been collected since 1841 but is no longer deemed a reliable indicator of ethnic origin with over 50% of ethnic minorities being born in the UK. Country of origin or ancestry was reported to be a reasonably reliable indicator but was problematic for mixed ethnic group. Name recognition software was reported to be a useful tool for identifying South Asians and could be used for validation purposes (Gill et al, 2006).

In summary, other indicators of ethnicity apart from ethnic group itself were country of birth, nationality, language spoken in the home, country of origin in conjunction with country of birth, skin colour (white and black), national/geographic origin, diet and religious beliefs. The UK Department of Health guidelines (2005) give many examples of good practice throughout the UK with a

variety of ethnicity indicators collected in practice, revealing the indicators collected were based upon local needs.

4.3.4 Methods of collection

Self-reported ethnicity was consistently recommended to be the gold standard and the reasons for this were discussed in many good practice guidelines and papers (Regenstein and Sickler, 2006; Commission for Racial Equality, 2002; Department of Health, 2005). As ethnicity is difficult to determine by observation, there is good evidence to justify the need for self-reported ethnicity data and active discouragement of obtaining ethnicity by observation. Observation may often lead to 'stereotyping' by skin colour and name and should be restricted to situations where obtaining self-reported data is not possible, e.g. language barriers. A USA paper by Hasnain-Wynia and Baker (2006), explained how staff should ask for these data and also emphasised the need for self-reported ethnicity. The HRET toolkit included training on how ethnicity data should be collected and again there was emphasis on the need for self-reported ethnicity (Health Research and Education Trust, 2007).

Results from hospital surveys conducted in the USA by the Robert Wood Johnson group revealed 61% of respondents reported usually asking the patient to self-identify. However, 25% filled in the ethnicity themselves based upon observation. They believed using the observation method was easier for both themselves and the patients as it avoided discomfort for both parties, they felt the data to be

accurate as they believed they knew their local population (Regenstein and Sickler, 2006).

As the literature has indicated, collecting self-reported ethnicity is important and this suggests it would also be informative to record the method of collection, i.e. self-reporting, observation or other, as important biases could occur if assumptions are made about the reporting (Commission for Racial Equality, 2002). Sugarman and Lawson (1993) revealed racial disparities varied according to the method of ethnicity data collection. The incidence of renal disease was reported to increase from 268 per million to 312 per million in American Indians/Alaska Natives after corrections to coding were made.

In the UK, the number of boxes an individual can select when providing ethnicity data is not restricted (Department of Health, 2005 and National Statistics, 2003). This was also reported to be the case in New Zealand where their training kit recommended patients should tick as many boxes as they felt necessary to describe their ethnicity (Gardi, 2003). However, this was not the case internationally where only a small proportion of countries allowed the selection of multiple boxes or provided specific categories for mixed race individuals (Mason et al, 2003).

There was evidence to show that self-reported ethnicity was the most effective way to obtain accurate information, especially for the 'mixed' or 'other' ethnic groups where methods, such as observation or name recognition were known to be poor indicators. Buescher et al (2005) reported discrepancies between published data on

racial classification and self-reported race in a study of 118,000 live births in North Carolina in 2002. Mothers chose to describe their ethnicity using over 600 different terms on birth certificates. However, two thirds of these mothers from Hispanic origin would have been coded as 'Other' using official ethnicity coding. This highlights the need to re-examine official categories, especially as ethnic and racial diversity continues to increase. It is also important to distinguish differences for missing data, occasions when staff 'failed to ask' and when patients 'preferred not to say' indicate two sets of problems (Department of Health, 2005).

Methods of ethnicity data collection other than self-report and observer assessment included the use of name recognition software. However, this method is becoming increasing unreliable, especially in younger generations where there the prevalence of mixed marriages is on the increase. However, use of such software can be useful for obtaining missing data or obtaining ethnicity where none was recorded. Chattar-Cora et al (2000) demonstrated successful use of obtaining ethnicity information from patient's medical notes. This method was used to identify most patients in the USA study, showing names can be used to obtain ethnicity data with some precision when no other data are available. Furthermore, the use of name recognition software in conjunction with country of birth data has been reported to result in improved accuracy of ethnicity (Sheth et al 1997, Swalen et al, 1997, Warnakulasuriya et al, 1999).

4.3.5 Barriers to collection

Site visits in the USA to six consortium member hospitals and a 1,000 nationwide hospital survey to focusing on current data collection practices found 30% of respondents reported problems collecting these data (Hasnain-Wynia et al, 2004). The barriers reported by Hasnain-Wynia et al (2004) were comparable to those found in the Robert Wood Johnson Report (Regenstein and Sickler, 2006); these being:

- Reluctance of staff to ask for this type of information for fear of offending the patient or encountering resistance
- Confusion about race/ethnicity categories
- Lack of understanding of the need to collect these data
- Reluctance of patients to provide this type of information
- Limitations of IT systems to capture this type of data
- Lack of resources (e.g. time constraints, staff training)
- Concerns that ethnicity data collection may expose the hospital to legal liability
- Lack of agreement of executive management on the need to collect these data

Many of the reports reviewed stated that there was a need to use the collected ethnicity data in research projects, for setting targets and comparing outcomes. The research should also be published to motivate healthcare professionals and patients to collect and provide these data having seen evidence of it being used for

meaningful research/changes/commissioning of policies to reduce health inequalities (Race for Health, 2006). Until ethnicity data are collected and reported and are shown to have a demonstrated use neither patients nor healthcare professionals will believe these data to be useful or important.

One of the other main barriers to ethnicity data collection was reported to be the patient perceptions. Patients would be more willing to provide data if the reasons for the data collection are explained; also healthcare professionals should appear comfortable when asking for these data (Baker et al, 2005). Baker et al (2007) reported 46% of patients worried that their data would be used by health care professionals to discriminate against them.

4.3.6 Interventions

All of the best practice guidelines agreed that the main intervention needed for improved completeness and accuracy of ethnicity data collection was staff training, followed by the provision of adequate resources for data collection, and thirdly 'use' of the data (Health Scotland, 2005; Department of Health, 2005; Commission for Racial Equality, 2002; Regenstein and Sickler, 2006; Health Research and Education Trust, 2007).

The UK DoH NHS 'practical guide to ethnic monitoring in the NHS and social care' (2005) described the importance of staff training and provision of information regarding the importance of ethnic monitoring, how the data will be used and the best methods by which to collect it. Training should be compulsory for healthcare

professionals that may be involved in collecting or using ethnicity data; for new members of staff this could be incorporated into induction training. Self-reporting was described as vital as ethnicity is part of an individuals' identity that may or may not be visible to a third party.

In the USA, the HRET toolkit provides a comprehensive training package for the collection of ethnicity data (Health Research and Education Trust, 2007). It has been written for multiple levels of healthcare professionals, including Chief Executive Officers, Clinicians, registration staff, database managers and patients, so that users can select the training information most relevant to them (Health Research and Education Trust, 2007). Once registered, the toolkit is free to access. The training emphasised the importance of the individual's perception of their own ethnicity, the need for self-reporting and a Patient Response Matrix which was a tool designed to help healthcare professionals respond to frequently asked questions in a standardised manner. The Patient Response Matrix lists common questions asked by actual patients and provides an appropriate standardised response. The Patient Response Matrix was designed to grow with the addition of information from users.

The 'Ethnic monitoring tool' developed by NHS Scotland is one of the most comprehensive training packages in the UK (Health Scotland, 2005). It was developed with NHS Scotland staff in mind and provides information relating to the reasons for carrying out ethnicity monitoring, who should be involved in monitoring and what needs to be in put place. Training materials provided included PowerPoint

presentations which can be downloaded and modified, train the trainer notes and role play scenarios are also available.

The importance of staff training was also emphasised by the Robert Wood Johnson Report where three different training methods were used across three separate hospitals. In Central Georgia, training in data collection was given to all new members of staff as part of their induction training program. In Boston, training on the importance of self-reporting and the importance of and uses of the data to improve services/healthcare was given to all staff involved in the collection of ethnicity data. In Seattle, members of staff working in the registration areas were quality reviewed; managers were able to identify members of staff recording a large number of unknowns or leaving the ethnicity field blank and provide training to these individuals to address the problems (Regenstein and Sickler, 2006).

4.3.7 Data quality and completeness

Incompleteness of ethnicity data is an on-going problem, reports based on incomplete or bad quality data can provide misleading and inconsistent results. In the USA, ethnicity data reported on death certificates was validated against data in the Indian Health Services database, 87% were found to be correct (Frost et al, 1994). Ethnicity data validated across three sources for AIDS patients showed 57% conflicts in American Indians and Alaskan Natives (Kelly *et al*, 1996).

Numerous studies, in the USA in particular, have compared self-reported data to official statistics and found inaccuracies (Buescher et al, 2005); it is important to

have more complete self-reported ethnicity information. Ethnicity data were assessed in 376 patients recently diagnosed with cancer, findings showed that medical records were well linked to self-defined ethnicity (Gotay and Holup, 2004).

The incompleteness of ethnicity data has been a major problem in the UK for cancer registration as up until recently they did not routinely collect these data but instead depended on obtaining these data from other sources. Jack et al (2006) reported only 23% of registry data having ethnicity compared to 81% in HES data, linking records was favoured as it helped reduce the duplication of work. In the USA, an initiative as part of the Surveillance, Epidemiology and End Results (SEER) program to assess the completeness of country of birth data reported only 67% to have recorded data with levels of completeness varying by ethnic group which suggested bias in the data collection (Lin et al, 2001).

The USA Centres for Disease Control (CDC) saw no improvement in the collection of 'race' data between 1994 and 1997 in the wake of the Revitalization Act of 1993 (Centres for Disease Control and Prevention, 1999). In 1995, it became UK government policy to collect ethnicity data in secondary care setting as an addition to HES. Completeness of ethnicity in HES has improved in London since the first collection in 1996/97 from 48% to 65% complete data in 2001/02 (HES online, 2004 and London Health Observatory, 2003).

The importance of data collection was increasingly recognised with new DoH initiatives put in place for GPs to collect ethnicity for newly registered patients

(Quality Outcomes Framework (QOF) points), although this initiative was unsuccessful and was abandoned. There is some way to go before databases have complete and self-validated ethnicity data. However, the Lambeth PCT project demonstrated that with dedicated resources, training and monitoring, improvements can be made in a relatively short amount of time.

4.4 Conclusion

To summarise, there is a need to increase awareness about the importance of routinely collecting ethnicity information. Self-reported ethnicity should preferably be collected as a mandatory item at the primary care level (with a 'not stated' option for patients who refuse to provide their ethnicity for coding purposes). 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 could also be collected at first hospital visit. Ideally databases could be linked between primary and secondary care so data are only collected once and verified at subsequent visits. One possibility would be to add ethnicity to the new patient Summary Care Record (NHS Care Records Service, 2011). Sangowawa and Bhopal (2000) showed successful collection of ethnicity data in a primary care feasibility study also demonstrating the ease with which ethnicity could be included on hospital referral letters by creating an automated field.

Many projects are on-going, e.g. the development of the NHS for Scotland toolkit. The Department of Health training developed in conjunction with the 2005

guidelines can be used to raise awareness and improve quality and completeness of ethnic data collection (Health Scotland, 2005 and Department of Health, 2005). At the cancer registration level, identification of high risk groups can only be based on the data collected. If these data are not available, incomplete and not validated, then reports based upon such data are unreliable.

4.5 Summary

This chapter reviewed literature discussing methods of improving ethnicity data collection. The next two chapters report investigations into the perceptions and experiences of ethnicity data collection from two different perspectives. Chapter Five presents results of focus groups with lay members from the South Asian community and Chapter Six the results from a survey of health care professionals.

Chapter 5 Focus groups with South Asian lay community members

5.1 Introduction

A component of the CRUK commissioned CanEth project was to explore barriers to ethnicity data collection by evaluating the perceptions and experiences of ethnic minority participants and their willingness to provide these data. This topic was investigated in the form of a series of five focus groups conducted with healthy lay volunteers from South Asian communities. South Asians were the largest ethnic minority group in 2001 making up 4.4% of the total England and Wales population and over 34% of the ethnic minority population (UK census 2001 [http://www.statistics.gov.uk/census2001/census2001.asp). In addition, South Asians from India, Pakistan, and Bangladesh are heterogeneous in terms of culture, language, religious beliefs, diet, migration history, educational attainment and social class despite having a similar outward appearance (Gill et al, 2007).

5.2 Methods

Five focus groups were conducted in collaboration with the Mary Seacole research centre at De Montfort University and the Ethnic Health Forum in Manchester through whom focus group facilitators were recruited. Facilitators spoke both English and either Mirpuri, Punjabi, Bengali or Urdu. Ethnic minorities selected for

focus groups were based upon the languages spoken by the facilitators. Discussions took place in the native language of each group. Two focus groups discussions took place in Urdu for males and females independently, three further discussions took place in Bengali, Mirpuri and Punjabi with the Punjabi group being the only mixed gender discussion. A topic guide, participant information sheet and consent forms were developed and approved by South Birmingham Local Research Ethics Committee as part of the CanEth project application (Appendix 1). Focus group facilitators used the topic guide which was developed to focus on five areas of interest (Table 5.1).The areas of interest and subtopics were developed in collaboration with the CanEth project team which included Mark Johnson who had expertise in the area of conducting focus groups with ethnic minority groups, and further refined following discussion with the CanEth Advisory Board. Lay South Asian volunteers were recruited by the trained facilitators from local sources such as, community centres and places of worship. The focus group discussions were also conducted at these venues to ensure the participants were in familiar surroundings in order to encourage open discussion. Facilitators aimed to recruit between five to ten participants per group and gender segregation was observed according to cultural customs for the Urdu, Bengali and Mirpuri speaking focus groups which were also conducted by a gender-matched facilitator.

Small incentives were offered to encourage participation. Facilitators selected the incentive they believed would be the most suited to their local population. The older Bengali group were provided with refreshments at the meeting and lunch

post discussion. The Urdu, Mirpuri and Punjabi groups were offered high street vouchers worth £20.

Table 5.1 Focus group topic guide

1. 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 these data?
2. 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?
3. 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
4. What about language, Religion, Culture: ▪ Do people ask, do you offer this information, do you mind. ▪ Are there problems with 'stereotypes' (Explain)
5. 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 these 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?

(Iqbal et al, 2008 and 2012a)

Discussions were recorded using either a cassette or digital recording device, then transcribed and translated by facilitators. Discussions ranged from 50 minutes in the smallest Punjabi group with five participants to 90 minutes in the Urdu females group with ten participants. Focus group discussions were quality checked by independent reviewers who listened to recordings in the native language and then reviewed the translated transcripts. The transcripts were reported to be accurate, with no key issues reported to be lost in translation. All groups apart from Urdu (males and females) were transcribed with the ability to identify participants by number. Participant number is provided in the results where possible in order to distinguish individuals dominating the conversation and to get a feel for the group dynamics.

The translated transcripts were analysed using a qualitative thematic approach. This process involved thorough examination of the data, comparing responses with one and another and identifying common themes. The focus groups undertaken were dependent on the availability of facilitators who were required to speak in the native tongue of the focus group and English, as follows:

1. Mirpuri speaking Muslim females of Azad Kashmiri origin
2. Bengali speaking Muslim males of Bangladeshi origin
3. Urdu speaking Muslim males of Pakistani origin
4. Urdu speaking Muslim females of Pakistani origin
5. Punjabi speaking Hindu males and females of Indian origin

Informed consent was taken by the facilitator where English was not the participants choice of language. A short introduction was given by the facilitator in the language appropriate for the group. Discussion took place in both the native language and English particularly where younger volunteers were involved e.g. Urdu females. I was present at two of the focus groups (Bengali males and Urdu females) and was introduced to the participants as the researcher and able to make additional notes.

5.3 Results

Five focus groups were conducted by trained facilitators, all speaking in the preferred language of their group and also in English. The number of participants in each group ranged from five to ten. The characteristics of the 36 volunteer sample are shown in Table 5.2.

Table 5.2 Characteristics of participants

			Gender			
Group	Country of origin	Language	M	F	Median age (range)	Total
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
Total			18	18	31.5 (18-70)	36

(Iqbal et al, 2008 and 2012a)

The facilitators used the topic guide which was specifically developed to focus on the five key areas of interest and consisted of sub-categories. The subthemes that

emerged in the focus group discussion within the main themes are summarised in table 5.3 and also highlight the order in which the result s are presented:

Table 5.3: Summary of focus group themes

	Main theme	Subtheme
1	General opinions on the collection of ethnicity	• Objections or worries about providing ethnicity information
2	Experiences of providing ethnicity information	• General feelings of providing ethnicity information • Effects of providing ethnicity information in healthcare situations
3	Ethnicity categories used in practice	• Thoughts on current ethnicity categories and preferred terms
4	Provision of data on language, Religion, Culture:	• Experience of being asked to provide data on language, religion and culture • Fears related to the risk of stereotyping
5	How *should* this information be collected?	• Recommended point of collection for routine ethnicity data capture (hospital/GP/other) • How ethnicity information should be collected and what staff should tell you

5.3.1 General opinions on the collection of ethnicity

Overall, participants thought accurate recording of ethnicity data was important. Most were proud of their origins and were familiar with the differences between their culture and other cultures, particularly other South Asians and displayed some understanding of the usefulness of ethnicity data in the healthcare setting. A number of participants also knew of the increased prevalence of certain diseases in minority ethnic groups and indicated this as the reasons why ethnicity data collection is important in healthcare:

"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]

"Sometimes certain illnesses are directly linked to our ethnicity. If a doctor does not know the right ethnicity he cannot do anything. For example stroke or diabetes is directly linked to ethnicity" [Urdu female participant]

"….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…." [Bengali male, participant 4]

Several participants mentioned the importance of monitoring access and uptake of services. Others mentioned the need for collection of ethnicity data for future planning. Younger participants felt providing ethnicity was acceptable in healthcare but not in other areas, such as job applications:

"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 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 participant]

"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 participant]

A small proportion (4 out of 36) did not understand the need for ethnicity data collection stating it was not relevant to treatment, or believed 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 participant]

Any objections/worries about providing this information?

When asked about objections or worries about providing ethnicity data the vast majority had no objections. However, most of the participants' experiences of providing ethnicity information was for job applications which was not viewed in a positively and many felt that it may discriminate against them getting the job. Several had concerns, or sometimes felt unease on occasions when the purpose of ethnicity data collection was not fully explained, and worried about being stereotyped. There was a feeling of dissatisfaction when the appropriate category did not appear on a form and feelings of the data not being utilised. One participant did not believe discrimination to be a problem given the multi-cultural nature of NHS work force:

"I feel uneasy sometimes and you start wondering why they ask me questions about my ethnicity" [Urdu male participant]

"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]

"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]

5.3.2 Experiences of providing ethnicity information

In general, when asked about how they felt about providing ethnicity information, the majority of people did not mind. Others only minded when they were asked at repeatedly. The majority wanted an explanation as to why the data were being collected and how it would be used.

General feelings of providing about your ethnicity information

"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 participant]

Positive experiences reported included a participant who did not speak English and was offered an interpreter which helped. The main reason given for negative experiences was inadequate ethnicity codes and the fact that they would be coded as 'other' which resulted in frustration and feelings of being insignificant:

"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]

"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]

No participants voiced objections to providing ethnicity information in a healthcare setting. However, there was some confusion regarding the procedure for ethnicity data collection in healthcare and the need for standardisation:

"My child was born in the same hospital yet they ask ethnic data about him whenever I took him to hospital" [Urdu male participant]

"Sometimes they ask these questions about ethnicity and sometimes they do not so we are not sure what is the standard routine" [Urdu male participant]

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]

"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 ethnic group categories used in practice, many participants felt additional ethnicity indicators, such as country of birth, language and religion should be collected in order to distinguish between South Asian groups. One participant also thought information on diet would be useful; another participant added that it would also be useful to be asked whether you were willing 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?

"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]

"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 ethnicity should not be confused with the colour of the skin" [Urdu female participant]

5.3.4 Language, religion and culture

All participants were happy to disclose their religion and language as long as they felt they were not being stereotyped. The discussion on culture centred on the reasons why religion was a better indicator of culture in South Asian communities than 'ethnic group'. Some Muslims participants felt they were stereotyped, particularly in connection with terrorism. Also, some Bengalis reported having experienced being called 'Paki' purely based on stereotypes relating to skin colour. Others did not feel stereotyping to be a problem and were proud of their language, religion and culture and did 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]

"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]

"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 participant]

"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 participant]

Do you think there is a risk of 'stereotyping'

"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]

"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]

"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]

"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 participant]

Stereotyping by healthcare professionals was an issue for some participants:

"Walk-in centres provide independent advice but I feel my GP knows my family history so makes assumptions about me" [Punjabi male, participant 3]

5.3.5 How information should be collected?

The Bengali focus group summed up how ethnicity data 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].

The majority of participants agreed that GPs should collect ethnicity data and that this information should then be made available to hospitals and other healthcare providers. There was a general consensus that not enough is known as to the value, importance and use of these 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]

"The information should be collected at the GP surgery as patients are already distressed in hospital" [Punjabi female, participant 1]

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]

"If they explain its importance before they ask question about ethnicity, I will feel more comfortable in providing such information" [Urdu male participant]

5.3.6 Discussion

On the whole, the majority of focus group participants were happy to provide ethnicity data in healthcare situations providing they were offered an explanation as why it was needed and how the data will be used. Focus groups participants felt

staff should seem comfortable when asking questions about ethnic origin. The exhibition of discomfort by staff could make patients suspicious of the purposes for the questions and exacerbate non-compliance. Baker et al (2007), trialed four rationale justifying the collection of race and ethnicity data in the USA (1) quality monitoring, 2) government recommendation, 3) needs assessment and 4) personal gain and reported changes in patient comfort levels. Levels were reported to be the highest when quality monitoring was stated as the reason for collection. Similar rationale could be explored in the UK based upon well-known artefacts such differences in disease burden or the promotion of equality in access to healthcare.

5.3.7 Discomfort

Several focus group participants had reservations about providing ethnicity data and conveyed feelings of discomfort when the purpose of the data collection and how it would be used was not fully explained. Participants considered a brief explanation of why the data were required and its intended use would increase the willingness to provide ethnicity. Insubstantial explanations, such as ‘it’s routine’ or ‘procedure’ or neglecting to offer any rationale were deemed unsatisfactory. It was felt that data collected ‘for statistical purposes’ were not utilised. These findings confirm previous reports from Pringle and Rothera (1996) and Hasnain-Wynia et al (2011) who recommended patients were informed of the reasons for collection and how the data will be used to improve services for patients.

5.3.8 Understanding of differences in disease risk and justification

Focus groups with lay members of South Asian communities overwhelmingly indicated their willingness to provide ethnicity data for healthcare purposes. A number of participants demonstrated an understanding of variations in disease patterns by ethnic group, particularly in the incidence of diabetes and heart disease. The increased risk of these conditions was highlighted as the main reason for accurate data collection.

5.3.9 Additional indicators

There was also a consensus that collection of additional items such as religion, language, country of birth and diet were required to distinguish between the UK's heterogeneous South Asian population. Focus group participants discussed descriptors of ethnicity they felt to be important in healthcare and also to distinguish between ethnic groups. Language, religion and country of birth were considered to be crucial for this group of South Asians who are culturally diverse despite having a similar outward appearance, and possessing similar genetic information.

In one focus group participants stated they would describe themselves as "British Muslims" excluding their country of origin as they felt their religious beliefs were

the most significant indicator of their culture, for example religion dictates diet and consumption of alcohol and tobacco and therefore could be linked to health risk.

5.3.10 Limitations of current categories

A selection of participants expressed frustration when their ethnic group was not listed on the form and they were forced to tick 'other'. The ethnicity categories were expanded from the original in both 2001 and 2011 based upon responses described in the 'other' category.

5.3.11 Streamlined collection

The focus groups also raised the issue of repeated ethnicity data collection and did not feel they should be asked for this information repeatedly, especially as it rarely changed. The majority agreed primary care was the preferred point of collection where patients are less distressed have some acquaintance with the staff. Raleigh (2008) also stated with 90% of all patient contact occurring in primary care there are more chances to capture this information in this setting (Raleigh, 2008). Several studies have reported the feasibility of automatically linking data collected in primary care to secondary care, eliminating the need for repetition (Pringle and Rothera, 1996; Information Services Division Scotland, 2010; Sangowawa and Bhopal, 2000).

Providing data at first hospital visit was viewed as acceptable but repeated collection at subsequent visits was not thought necessary, however repeat visits

were a good opportunity to verify recorded data. Recent initiatives such as NHS electronic Summary Care Record which enable the sharing of selected patient information should prove useful for healthcare professionals and reduce treatment delays, but should also ease the problem of repeatedly giving the same information for patients; ethnicity could be incorporated as part of the patient demographics data (NHS Care Record Service, 2011).

5.3.12 Limitations

The focus groups conducted for this thesis were limited to South Asian groups. However, South Asians were the largest ethnic minority group in the UK 2001 and 2011 censuses. Focus groups with Bengali females and Mirpuri males were missed due to a lack of finding an appropriate facilitator. Younger participants were bilingual (e.g. in the Urdu females and Punjabi groups) which enabled me to gather data from English speaking and non-English speaking South Asians, although the discussions took place mainly in the native language in order to include all participants. Focus groups with English speaking ethnic minority groups such as Black African, Black Caribbean, White Irish and Mixed populations would have been informative but were not conducted due to time limitations and financial constraints imposed by the CanEth feasibility study.

A further limitation to be considered is that I was not present at all the focus groups, therefore did not have firsthand experience of all the discussions. Moreover, the sample was relatively young with a median age of 31.5 years and

would have benefitted from including older ethnic minorities who are more likely to be health service users. Selection bias was also likely to be an issue as the focus groups were made up of a purposive sample and all participants were voluntary. As a consequence, these groups may be biased in favour of providing ethnicity data and the results may not be as generalisable to other South Asians.

5.4 Conclusion

In summary, results showed ethnicity should ideally only be collected once by GP or at first hospital visit and verified at subsequent visits if needed. 'Ethnicity' information collected should include religion, country of birth and language to account for cultural differences. In general, there was no objection to providing ethnicity data for healthcare purposes. There was some understanding of differences in disease patterns, e.g. higher incidence of diabetes in South Asians, and the importance of recording these trends. Explanations of why the data is needed and how it will be used to improve health care/services would increase willingness to provide the data. Ethnicity data should be collected at GP practice level where the staff are familiar and the patients less distressed than in a hospital setting.

To follow on, Chapter Six now presents the results of a survey of healthcare professional perceptions and experiences of collecting ethnicity data.

Chapter 6 Health care professionals ethnicity data collection survey

Ethnicity data are generally collected once in primary care at the point of first registration. In the hospital setting, in both in- and outpatient clinics ethnicity is usually requested by reception staff when a patient arrives for an appointment. However, in cases where the patient presents in an accident and emergency situation this information may be requested from relatives or friends accompanying the patient.

Despite the drive towards improving UK ethnicity data collection in healthcare, relatively little is known about how healthcare professionals (HCP) in the UK perceive the collection of these data. The systematic literature review of ethnicity data collection methodology reported in Chapter Four revealed healthcare professionals' perceptions to be a major barrier to ethnicity data collection. Research into healthcare professionals attitudes towards ethnicity data collection have been reported in primary care where it was generally viewed in a positive light (Pringle and Rothera 1996). Sixteen GPs and practice managers described ethnicity data collection to be valuable for service evaluation and health promotion purposes (Sangowawa and Bhopal, 2000). In spite of this, ethnicity data collection has been shown to be poorly collected in primary care with research concentrating on

acceptability, resources, database restrictions and ethnicity coding (Kumarapeli et al, 2006, Pringle and Rothera, 1996).

'*Barriers to collection*' was as one of seven themes identified by the literature review of ethnicity data collection reported in Chapter Four. HCP barriers to ethnicity data collection in two reports from the USA included a fear of causing offence to patients, meeting with resistance, confusion with regards to ethnicity categories and a lack of understanding of the need for the data (Hasnain-Wynia et al. 2004, Regenstein and Sickler, 2006). Yet, little is known about healthcare professional perceptions and experiences of ethnicity data collection in secondary care in the UK.

The aim of this chapter was to explore the likely reasons for gaps in ethnicity data by evaluating the perceptions and experiences of UK HCPs responsible for collecting this information.

6.1 Methods

The questionnaire was based on one previously developed by the Centre for Evidence in Ethnicity Health and Diversity (CEEHD) at Warwick and was adjusted by the CanEth working group. The modified two page questionnaire consisted of nine items including: 1) perceived importance, 2) current practice, 3) reasons for not collecting ethnicity data, 4) problems encountered when data are collected, 5) disease areas, 6) method of collection, 7) items collected, 8) use of name recognition software and 9) ethnicity data collection training (Appendix 2). The

questionnaire was intended to be quick and easy to complete, ensuring most questions could be answered using tick box responses but also allowed additional space for comments. The questionnaire could be printed, completed and returned by post or emailed.

The questionnaire was aimed at clinicians, managers and nurses and any other healthcare professionals 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)
- 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 Specialist Library for Ethnicity and Health (SLEH) website (http://www.library.nhs.uk/ethnicity/)
- Thread created on Academic Clinical Oncology and Radiobiology Research Network (ACCORN) thread (http://www.acorrn.org/ResearchDB/)
- Thread on NHS discussion forum and news item on `new@networks' electronic bulletin in June (http://www.networks.nhs.uk/forums/)
- News item in the Welsh Cancer network 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/)

The questionnaire was distributed throughout England and Wales in 2007, via Minority-Ethnic-Health and ALLSTAT JISCmail lists (National Academic Mailing List Service for Academic and research communities). In addition, the questionnaire was also circulated to the 23 Race for Health Primary Care Trusts and to registered members of the Race for Health mailing list. It was posted on the Warwick CEEHD website and a link was placed on the NHS Evidence - Ethnicity and Health (formerly the Specialist Library for Ethnicity and Health) website. The questionnaire was circulated to the National Cancer Research Network (NCRN) head office and 24 Cancer Networks in England and Wales. Threads on relevant websites, such as, NHS and ACCORN discussion forums were created. There was a special interest in the cancer networks because the survey formed part of the CanEth CRUK commissioned project.

A four week deadline to return the questionnaire was set and extended for a further four 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.

The role of healthcare professionals are summarised in tabular form. Responses to questionnaires are presented as charts or direct quotes in the case of open ended

questions. Quotes are presented as anonymous as possible without losing information, respondents role and geographic area were provided where available.

6.2 Results

Thirty healthcare professionals completed and returned the questionnaire. The sample was distributed throughout England & Wales, with eight from Midlands, six from Wales, eight from the North of England, two from South England whilst six responders did not provide location information. Breakdown of respondent by role in the NHS is shown in Table 6.1, the majority of questionnaires been completed by clinicians, nurses or information officers. Role was not provided for six respondents.

Table 6.1 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
Total	**30**

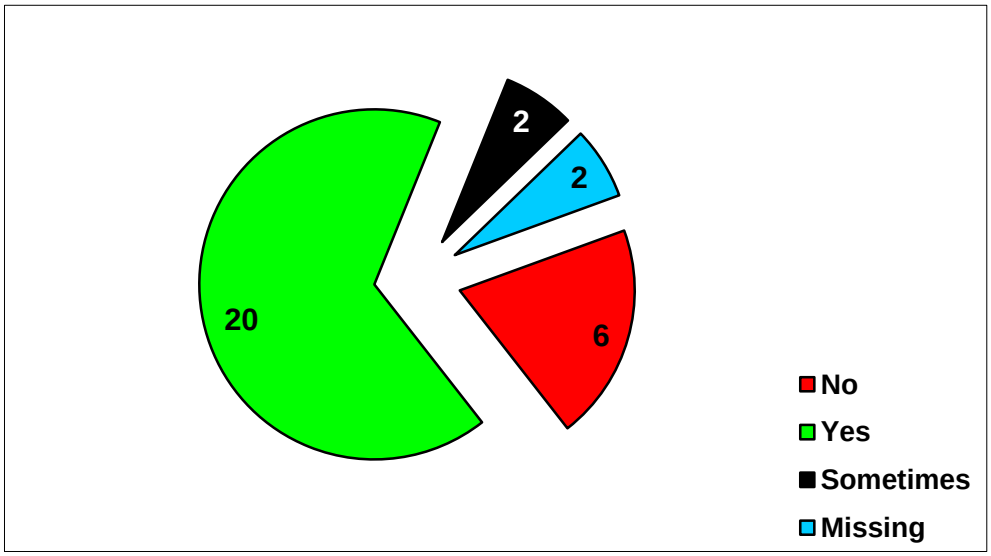
(Iqbal et al, 2008)

6.2.1 Do you attempt to collect any ethnicity data on patients?

Two-thirds of respondents (n=20) attempted to collect some ethnicity data; two further respondents (6.5%) did not consistently collect ethnicity data, six (20%)

reported not collecting ethnicity and two (6.5%) did not complete the question, see Figure 6.1.

Figure 6.6.1: Attempts to collect data



(Iqbal et al, 2008)

6.2.2 Reasons given why ethnicity data are not collected

Of the 30 responders, 20 routinely collected ethnicity and 10 did not. Reasons for not collecting ethnicity were requested and are presented in Table 6.2.

Table 6.2 Reasons 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" (Cancer professional, Wales)
"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)
"We have not to date regarded it as sufficiently important" (Consultant, Wales)
"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)
"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)
"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)
"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)
"Ethnicity data is not part of the datasets that are collected" (Information Manager, Yorkshire)
"Not part of my job" (Radiographer, Gloucestershire)
"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)
"Our data collection is poorly resourced as it is so we have to stay entirely focused on what is clinically relevant" (Oncologist, Birmingham)

(*Note: Responders identity presented as anonymous as possible without losing information). (Iqbal et al, 2008)

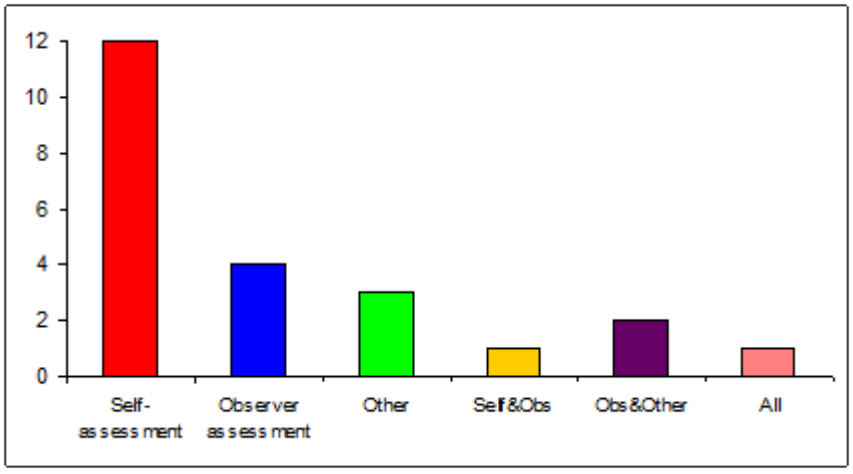
6.2.3 Disease areas for which ethnicity data are routinely collected

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

6.2.4 Method of collection

The majority of respondents who collected ethnicity data reported using the self-report method (n=12), assessment by observation was used less frequently (n=4). A number of respondents reported using a combination of methods, e.g. self and observer assessment (Figure 6.2).

Figure 6.2 Data collection methods (n=22)

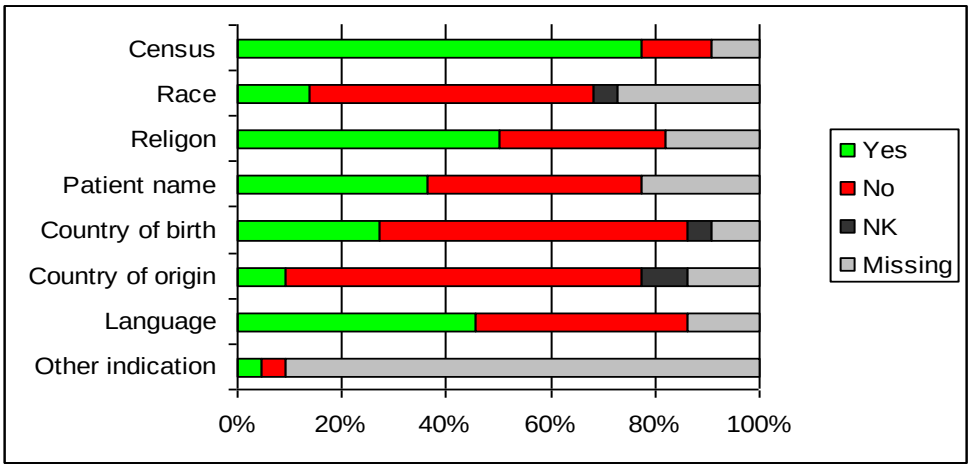


(Iqbal et al, 2008 and 2012b)

6.2.5 Indicators routinely collected

Ethnicity data was most commonly reported to be collected using the 16 census categories. Other routinely collected ethnicity indicators included religion, language and country of origin. Race and country of birth were the least likely to be collected. 'Need for an interpreter' was listed as an 'other' indicator (Figure 6.3).

Figure 6.3 Which indicators of ethnicity do you routinely collect?



(Iqbal et al, 2008 and 2012b)

6.2.6 Problems encountered during ethnicity data collection

The 22 patients who collected ethnicity were asked if they had encountered any problems in the process of collecting these data and were asked to provide details. Respondents experienced a variety of problems, see Table 6.3.

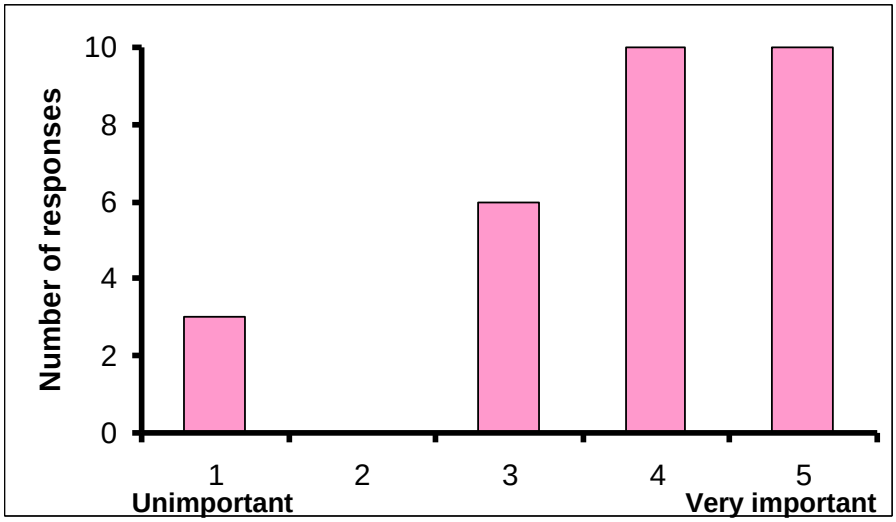
Table 6.3 Problems encountered collecting ethnicity data

"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)
"Patients 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). (Iqbal et al, 2008)

6.2.7 Perceived importance of ethnicity data collection

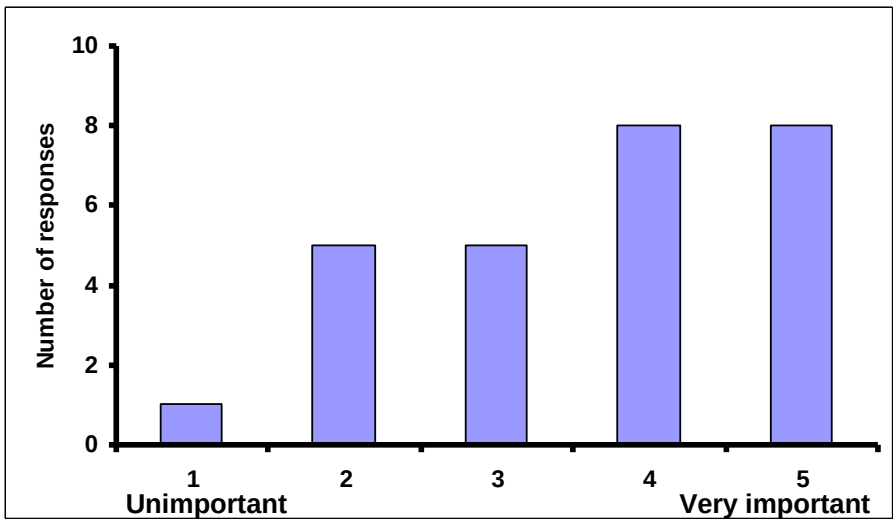
Figure 6.4 Please rate how important you personally think the collection of ethnicity data is?



(Iqbal et al, 2008)

The majority of respondents believed the collection of ethnicity data was more important to them personally compared to their perception of its value to their organisation. This may be symptomatic of the organisations weak policies on ethnicity data collection and lack of training provision (Figure 6.4 and 6.5).

Figure 6.5 Please rate the value of collecting ethnicity to your organisation



6.3 Discussion

A number of issues were identified by the HCP survey, including the methods of collection utilised. Self-report is universally agreed to be the best method of collection and recommended by many guidelines as the ideal (Commission for Racial Equality, 2002; Department of Health, 2005b; Regenstein and Sickler, 2006). Yet, assessment by observation alone judged by appearance (such as skin colour, hair colour and/or type) although discouraged by guidelines was the second most used technique. Explanations given for using observation included the avoidance of discomfort and confrontation, but also the fear of causing offence to patients.

An earlier project focusing on the collection of ethnicity data in the NHS also identified a fear of offending patients as a barrier to collection. This was accompanied by a fear of being accused of discrimination, concern that the questions were too sensitive and embarrassment when asking the questions (Johnson et al. 1993). Other concerns revealed by the survey included dealing with younger generations who are more likely to be British born and may wish to identify themselves as British. One respondent stressed the difficulty of obtaining self-reported data where the patient is unwilling to provide ethnicity (Department of Health, 2005a). The Commission for Racial Equality (2002) discourages other methods as a rule but stipulates they may be used in circumstances where self-report is not possible. In surveys carried out by the USA Robert Wood Johnson Foundation 61% of healthcare professionals reported using the self-report to obtain ethnicity, but 25% utilised the observer method. The healthcare professionals felt

the recorded data to be accurate based upon their knowledge of the local population, and believed that this eradicated discomfort for both themselves and patients (Regenstein and Sickler, 2006).

Several HCP survey respondents reported not collecting any ethnicity data. For many, this was a consequence of their, or their organisation's lack of awareness of the importance of the data and believed it not to be relevant to patient care or treatment. Exceptions to this were if the information was required for a specific purpose, such as meals, requesting an interpreter or for clinical trial participants (where requested). Physicians from the USA felt ethnicity and race to be clinically irrelevant (Hasnain-Wynia et al, 2010). Other barriers included worries about privacy and the legality of collection coupled with discomfort and resistance for patients and staff. These findings concur with previous reports by Regenstein and Sickler (2006) who reported the main barrier to data collection is staff not knowing why the data is of importance. This was not a major issue in the HCP survey where only one participant stated ethnicity was 'not relevant to care or treatment'.

Several respondents reported difficulties in releasing staff for training to attend off-site courses. However, training packages which offer a range of free material online have been developed by Health Scotland and Health Research and Education Trust which can be used for in-house training (Health Scotland, 2005; Health Research and Education Trust, 2007. Weinick *et al* (2007) reported releasing a few members for 'train the trainer' sessions to be a viable alternative.

A selection of respondents in the survey of healthcare professionals reported a lack of understanding of the need for ethnicity data, a lack of training on how to collect it along with a lack of resources. The source of this problem is ultimately the lack of training received by healthcare professionals. Training should include giving explanations to patients as to the reason for the collection and how the data will be used. Standardising the phrasing of questions, and the method and point of collection, available responses and explanations to frequently asked questions as suggested by Hasnain-Wynia et al (2004) may be beneficial to both healthcare professionals requesting these data.

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.

6.3.1 Limitations

The survey of healthcare professionals was limited in time and resources which meant I was not able to carry out large mailshots. Instead links to the questionnaire were posted on websites, forums, newsletters and mailing lists and a small mailshot to the NCRN. Unfortunately, the methods used meant I was unable to calculate a response rate. As it was, only 30 completed questionnaires were received despite

extending the deadline for returns. In hindsight, an online questionnaire may have been easier to complete and resulted in increased responses.

Circulation of the questionnaire amongst groups with an active interest in ethnic minority health (e.g. Race for Health and Ethnic Minority Health mailing lists) may have biased the findings of the survey in favour of the collection of ethnicity data. In hindsight and given more time and resources, I could have targeted mailshots at other groups e.g. primary care networks. Based upon the findings of my systematic literature review of ethnicity data collection, surveying reception staff may have provided a more realistic account of the barriers experienced on a day to day basis and how these vary according to the environment e.g. between primary care and secondary care. However, given the paucity of research in this area the data provides a useful insight into the perceptions and experiences of healthcare professionals and provides a basis for further work.

6.3.2 Summary

In the introduction I discussed the under-representation of ethnic minorities in clinical trials and the problem of poor ethnicity data collection. The previous three chapters have revealed much can be done to improve the current state of collection primarily by using available data to show how it can be used to improve health care provision and to highlight the problems with the data. This may motivate the collection and provision of data for use in future reports. Secondly, the provision of training to empower health care professionals to collect these data

and how to collect it, e.g. the importance of self-reported ethnicity would be beneficial.

In the next two chapters, I delve into ethnic minority participation in RCTs. Commencing with a systematic literature review of interventions used to facilitate the recruitment and retention of ethnic minorities into clinical trials in Chapter Seven. This is followed by an exploration of Hospital Episode Statistics (HES) data from a large teaching hospital and trial recruitment data from a cancer research network in Chapter Eight to assess the quality and completeness of ethnicity data collected locally. I then proceed to use these data to assess the representativeness of ethnic minorities in cancer clinical trials. Chapter Nine describes my journey towards establishing ethnicity data collection in conjunction with a CRN and uses these data to explore barriers to participation.

Chapter 7 Systematic literature review 2: Interventions to improve the recruitment and retention of ethnic minority patients into RCTs

7.1 Aim of the review

The aim of this chapter is to scrutinise the evidence base and synthesise the available information to identify effective interventions that have been used to facilitate the recruitment and/or retention of ethnic minority populations into RCTs.

7.2 Background

Ethnic minorities are reported to be under-represented in clinical trials. This topic will be explored in depth in Chapter Eight where I will attempt to investigate whether this is a problem using several locally sourced datasets. There has been some research into barriers to and attitudes towards participation in clinical trials suggesting clinicians may be hesitant to offer ethnic minority patients clinical trials due to time restrictions and a lack of resources (Ross et al, 1999, Ward et al, 2000). Hussain-Gambles et al (2004) reported results from an exploration of issues surrounding the participation of British South Asians in clinical trials. Common

deterrents to participation were concerns about drug side-effects, busy lifestyles, language, mistrust and not feeling part of the British society.

Two years later, Fayter (2006) published the results of a systematic review of barriers, modifiers and benefits of participating in cancer clinical trials. Fayter found the evidence was too weak to permit a clear interpretation. To improve the participation rates of ethnic minority patients into clinical trials we must identify methods which effectively overcome these barriers to participation as well as to continue to identify deterrents in the Black, Minority and Ethnic groups.

7.3 What's already known

Before I started this systematic review I searched the Database of Abstracts of Reviews of Effects (DARE) and the Cochrane Database of Systematic Reviews (CDSR) to ensure this review was not already being conducted elsewhere. Two systematic reviews in similar areas had been completed. Ford et al (2007) conducted a literature review of barriers to recruiting underrepresented populations to cancer trials. McDaid et al (2006) reviewed the literature with a view to identifying interventions to increase participation of cancer patients in RCTs. However, this review included all patients and was not specific to ethnic minorities and both reviews were only interested in cancer trials. Both found a dearth of research in this area. McDaid et al, found only eight articles meeting the inclusion criteria whilst Ford et al found five.

Both of the above identified reviews were restricted to cancer and only one considered interventions to improve participation in RCTs. Expanding to trials outside of cancer may yield more results. In addition, interventions that may be effective in a cancer population may not be so effective in other areas since the word 'cancer' provokes fear in people. Expanding the review to all clinical trials may also permit comparison by disease type. Furthermore, the effectiveness of interventions may also differ by ethnic group concerned, and this issue has not been addressed to date.

7.4 Focus of the current systematic literature review

My review focused upon interventions to improve the recruitment and retention of ethnic minorities into clinical trials. All disease sites were considered as well as all ethnic minorities. All studies with the exception of screening studies were considered. Screening studies were excluded because it was hypothesised that there are inherent differences in the process of deciding to participate in research between individuals with and without specific health problems. Joining a screening study where a 'healthy' individual may receive bad news may not be welcomed.

7.5 Methods

This systematic review was both conducted and reported according to the Centre for Reviews and Dissemination's guidance for undertaking reviews in health care,

University of York (NHS Centre for Reviews and Dissemination/University of York, 2001).

7.5.1 Search strategy

The following steps were undertaken in order to create the search strategy:

1. The research question was broken down into component parts/concepts
2. A list of search terms for each concept was created through the use of synonyms, abbreviations, differences in spelling (transatlantic differences), identification of related terms
3. Imploding and exploding of Medical Subject Headings (MeSH) and exploration of indexed subject trees helped develop the concepts and the list of search terms
4. Use of truncation and wildcards ($)
5. Relevant search terms for each concept were combined using operators: OR, AND, NOT
6. Proximity of search terms to one another was also utilised, e.g. adj5 displays terms within five words of another pre-specified term
7. Limits applied included articles written in English, restricted to humans and articles published between 1990 to Current (Nov 2012 at the time of review)

All randomised clinical trials, pilot studies and feasibility studies were included. The systematic literature review search strategy focused on four key concepts:

Key concept 1: Ethnicity

Ethnicity were taken from the previous literature review reported in Chapter Four, see Table 4.1.

Key concept 2: Recruitment and retention

Expanded terms:

- Patient Selection
- Patient Participation

Mapped terms and keywords:

- Recruit$ or enrol$ or register$ or screen$
- Retention or retain$
- Participa$ or inclu$
- Enter$ or nonentry or non-entry or non entry

Key concept 3: Interventions

Expanded terms: None

Mapped terms and keywords:

- Intervention$
- Strateg$
- Initiative$

Key concept 4: Randomised Clinical Trials

Expanded terms:

- Randomized Controlled Trials as Topic
- Clinical Trial
- Pilot Projects/ or pilot stud$.mp.

Mapped terms and keywords:

- Randomised Controlled Trial$ or RCT$.mp.

The search terms used for 'ethnicity' are based upon the previous systematic literature review of ethnicity data collection developed in conjunction with the Centre for Evidence in Ethnicity, Health and Diversity (http://www2.warwick.ac.uk/fac/med/research/csri/ethnicityhealth/).

Literature searches were conducted in Embase, MEDLINE, PsycINFO CINAHL, AMED, ASSIA and Race Relations databases using Ovid, EBSCO and Proquest platforms/engines. The literature search was limited from 1990 to the time of search (Nov 2012) with the view to capture literature pre- and post the Race Relations Act amendment in 2000 and the USA National Institute of Health (NIH) Revitalisation Act in 1993. The search was also limited to articles and reports written in the English language due to the time limitations of this thesis.

7.5.2 Inclusion and exclusion criteria

Populations of interest were Black, Minority and Ethnic groups as specified in Table 7.1. Articles reporting on ethnic minority populations within their countries of origin were excluded. All interventions, strategies, initiatives, incentives or techniques not used to facilitate recruitment or retention were excluded. Articles not comparing recruitment and/or retention by intervention and/or ethnicity minorities were also excluded. All trials designs were considered with the exception of screening trials and retrospective comparisons.

Table 7.1 Selection criteria and inclusion and exclusion criteria

Selection criteria	Inclusion criteria	Exclusion criteria
Population	Patients from a Black, Minority or Ethnic group. This includes indigenous populations of invaded countries e.g. Australia and USA OR Under-represented/under-served populations	Black, minority and ethnic groups in trials conducted in their countries of origin
Interventions	Any intervention or strategy or incentive or initiative OR any other device/technique etc used to facilitate recruitment into an RCT	Results of a trialled: Intervention Strategy Incentive Initiative etc that is NOT focused on recruiting to an RCT
Outcomes	Rates of ethnic minorities recruited/enrolled/registered into a trial as a result of an applied intervention.	Comparison by strategy not presented
Study design	Any of the below providing they are discussing recruiting to an RCT: Clinical trial Randomised Clinical Trial Pilot study Feasibility study	Screening trials Retrospective comparisons Studies with no comparison data

7.5.3 The selection process

The review process was broken down into three screening stages: 1) title, 2) abstract and 3) article review. Articles deemed relevant at stage three were subjected to quality assessment and data extraction. Three main steps were taken in order to minimise selection bias. Firstly, two reviewers independently screened titles, abstracts and articles. Any differences were resolved by consensus or by referral to a third reviewer. Secondly, two reviewers independently applied the inclusion criteria to the full articles. Again any differences were resolved by consensus or referral to a third reviewer. Thirdly, two reviewers' independently extracted data from the final selection and any differences were resolved as previously described.

I was the main reviewer and was assisted through the review process by both supervisors. Margaret Thorogood acted as the second reviewer, whilst Janet Dunn acted as the third reviewer when required.

7.5.4 Data extraction

The data extraction form was designed to capture the following information:

1. Target population
2. Intervention or strategy
3. Study design
4. Outcomes measured or reported

5. Setting in which they were conducted
6. Country of origin of the study and year of intervention
7. Numbers approached or screened, and numbers recruited and retained
8. Comparison between strategies and/or ethnic groups
9. Details of before and after comparisons where appropriate
10. Any other information

Themes identified during the course of the review were included as part of the data extraction framework as follows:

- Use of incentives, financial and non-financial (provision of free healthcare, transport to and from study visits, provision of refreshments or meals, gifts, money or gift vouchers). These also differed as to whether they were offered to potential participants or recruiters
- Involvement of community members or community organisations, e.g. places of worship
- Language (study materials, use of bilingual recruiters and similar)

7.5.5 Quality assessment

Quality assessment was performed on the final selection of articles in this review which focused on RCTs. Studies were quality assessed in accordance with the Centre for Reviews and Dissemination (CRD) guidelines for '*Undertaking Systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews'* (University of York, 2001, see Table 7.2). The quality

assessment criteria included whether study investigators had taken steps to minimise selection, attrition and performance bias and contamination. Further to this, whether the interventions and target populations were clearly defined was assessed. Quality assessment was conducted by two reviewers independently and presented according to the design of the study.

Table 7.2 Assessment of study quality

Selection bias (RCTs only)	
Was the method used to assign patients really random?	Yes/no/unclear
Was the allocation to the intervention concealed?	Yes/no/unclear
Selection bias (uncontrolled/before and after studies)	
Retrospective or prospective study?	
Was the patient selection process described?	Yes/no
Were details provided of the population from which the sample was selected?	Yes/no
Were there inclusion criteria?	Yes/no/unclear
Were all eligible patients invited to participate?	Yes/no/unclear
Is it possible that the investigators had discretion over who was selected?	Yes/no/unclear
Attrition bias (all studies)	
Were at least 80% of patients considered at follow-up?	Yes/no/unclear
Was it similar across groups?	Yes/no/unclear
Was a valid ITT analysis carried out? (all eligible patients included in the analyses)	Yes/no/unclear
The intervention (all studies)	
Did the design protect against contamination?	Yes/no/unclear
Did the design protect against performance bias?	Yes/no/unclear
Further comments:	
Relevance (all studies)	
Was the nature of the intervention clear?	Yes/no/partially
Was the target of the intervention clearly defined?	Yes/no/partially
General comments on relevance/ applicability	

7.5.6 Data synthesis

Key characteristics of included studies are presented in addition to a narrative synthesis of review findings. A table of excluded studies from the article screening stage and the reason for their expulsion are presented in a separate table within the results section.

7.6 Results

A total of 2643 hits were obtained but reduced to 2127 following de-duplication. Hits from each database with and without duplicates are shown in Table 7.3.

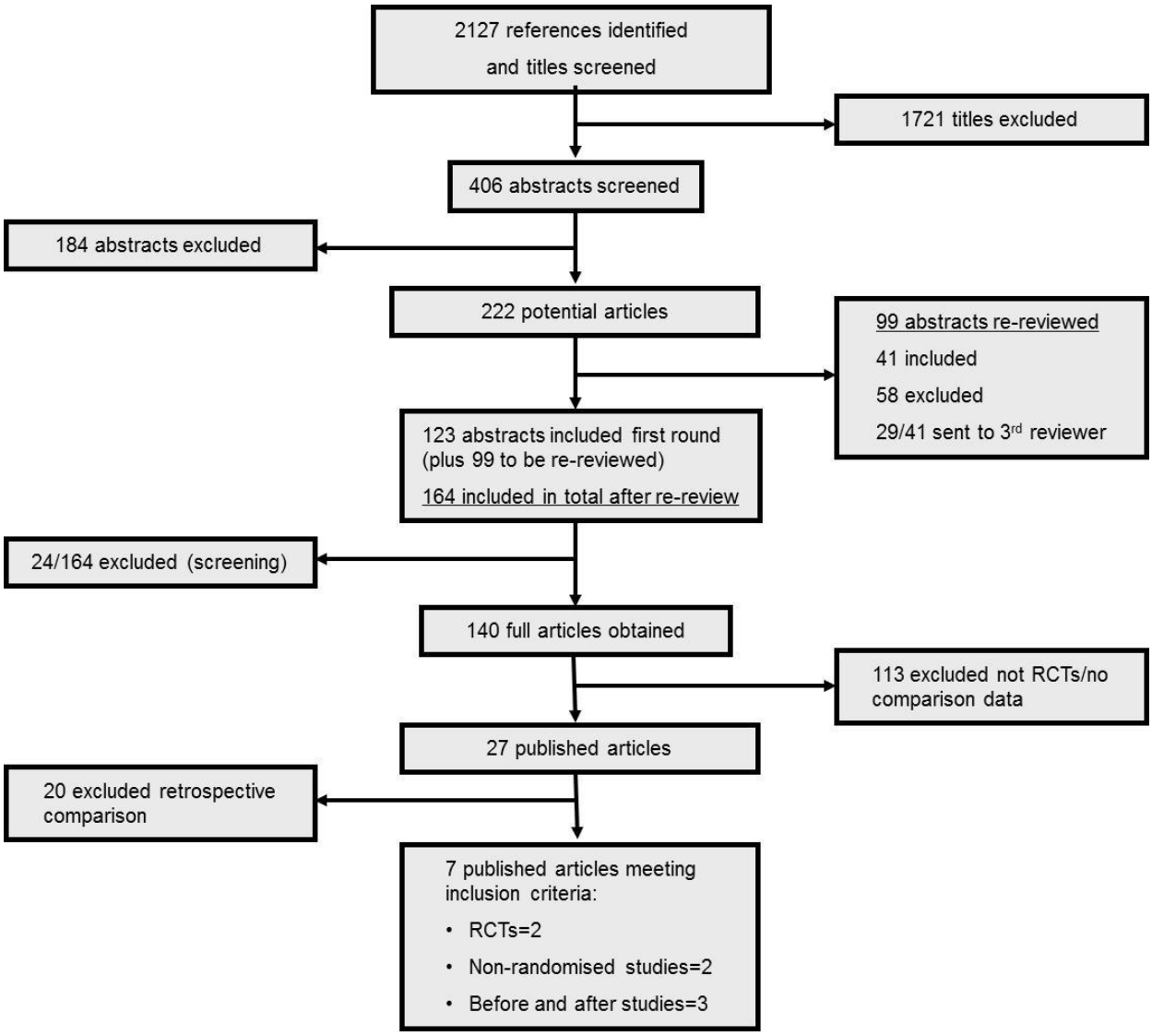
Table 7.3 Results of search by database

Host	Database	Hits	Hits without duplicates
Ovid	Medline	1510	1455
Ovid	Embase	687	389
EBSCO	Cinahl	262	150
Ovid	Amed	75	75
ProQUEST	PsychINFO	96	46
EBSCO	Race Relations	13	12
	Total	2643	**2127**

7.6.1 Study selection

The search identified 2127 references for screening (Figure 1). Of these, 406 abstracts were reviewed and 99 of these were subjected to re-review following disagreements in classification between the two reviewers. One hundred and sixty four articles were selected for article review, however, 24 articles were excluded because they focused on screening programs. Full articles for 140 references meeting the inclusion criteria were obtained for detailed review. Seven of the 140 articles met the inclusion criteria. Many of the articles excluded were retrospective comparisons. The study design was varied in the seven selected studies, two were RCTs (Du et al, 2009, Kiernan et al, 2000), two non-randomised comparisons (Cook et al, 2010, Horowitz et al, 2009) and three were before and after comparisons (Harris et al, 2003; Germino et al, 2011; Martin et al, 2011).

Figure 7.1 Process of study selection



7.6.2 Excluded studies

The majority of papers were excluded at abstract review because they did not provide any comparison with other recruitment strategies. However, the majority of papers excluded at the article review stage were due to the nature of the evidence, with many articles describing the lessons they learnt during the course of their research or reporting retrospective recruitment figures for a single group, and

therefore not providing comparison data. Other articles were excluded because the interventions were not focussed on facilitating recruitment to an RCT. Table 7.4 summarises reasons for exclusion at the article stage.

Table 7.4 Description of excluded articles (continued on the next page)

First author and year of publication	Description	Reason excluded
Anderson 2007	No comparison as such. Recruitment of African American was slower than non-African American using conventional methods therefore additional/tailored strategies were used	Exclude as there is not a true comparison. The African American tailored strategies were introduced part way through once they realised there was a problem so timespans different
Arean 2003	Methods of recruitment in 2 trials	Two separate trials, no direct comparison data
Bistrick 2010	Data presented by recruitment methods and ethnic group	Retrospective study, rejected due to lack of/poor comparison data
Douglas 2011	Recruitment methods, same target group (South Asians)	Emphasis of the article was on the screening stage of the trial not recruitment
Ellish 2009	Comparison of recruitment methods, same target group (African American)	Rates not accurate as numbers attending health fairs was estimated as was the number of people approached directly
Fitzgibbon 1998	Two different trials (Hip Hop vs FRITAA). Different strategies and different populations (hip hop included children)	Exclude because no like with like comparison. Hip Hop low income community and children. FRITAA medium level income adults
Gallagher 2004	Three recruitment source categories by ethnicity (Caucasian and Latino)	Retrospective study, rejected due to lack of/poor comparison data
Gallagher Thompson 2006	Three recruitment source categories by ethnicity (Caucasian and Chinese)	Retrospective study, rejected due to lack of/poor comparison data
Hatfield 2010	Comparison of several recruitment methods by ethnic group	Retrospective study, rejected due to lack of/poor comparison data

Heiney 2010	Implementation of the Heiney-Adams recruitment framework. Comparison between strategies	Retrospective study, rejected due to lack of/poor comparison data
Kennedy 2005	Recruitment techniques: mass mailing, mass media and personal contacts	Retrospective study, rejected due to lack of/poor comparison data
Kennedy 2010	Comparison by strategy and ethnicity (African American vs Non-Hispanic whites)	Retrospective study, rejected due to lack of/poor comparison data
Lindenstruth 2006	Comparison of various methods by ethnic group	Retrospective study, rejected due to lack of/poor comparison data
Mhurchu 2009	Comparison of various methods by ethnic group	Retrospective study, rejected due to lack of/poor comparison data
Sharp 2008	Comparison of flyers vs approaching folks on the street	Retrospective study, rejected due to lack of/poor comparison data
Stoy 1995	Comparison of various methods by ethnic group (Whites vs Blacks)	Retrospective study, rejected due to lack of/poor comparison data
Sweet 2008	Comparison of various methods by ethnic group (Whites vs Blacks)	Retrospective study, rejected due to lack of/poor comparison data
Unson 2004	Comparison of various methods by ethnic group (whites, African American, Hispanics)	Retrospective study, rejected due to lack of/poor comparison data
Vollmer 1998	Recruitment/retention by ethnic group but not by strategy	Retrospective study, rejected due to lack of/poor comparison data
Wisdom 2002	Compares different methods of recruiting African American	Retrospective study, rejected due to lack of/poor comparison data

7.6.3 Characteristics of included studies

Of the seven studies included in the review, three focused on recruiting patients to cancer treatment trials (Cook et al, 2010; Du et al, 2009; Germino et al, 2011), two on diabetes treatment trials (Horowitz et al, 2009 and Martin et al, 2011), one on a heart disease treatment trial (Kiernan et al, 2000) and one on a smoking cessation trial (Harris et al, 2003) see Table 7.5. Three studies focused on one ethnic minority group (Harris et al, 2003; Kiernan et al, 2000; Martin et al, 2011), while three studies included multiple groups alongside the ethnic minority group of interest (Cook et al, 2010; Germino et al, 2011 and Horowitz et al, 2009), and one targeted Whites only alongside the ethnic minorities group of interest (Du et al, 2009).

The numbers of participants per study varied widely from 33 to 8532. The studies were conducted between 1999 and 2007 (two did not provide dates). Three focused on one intervention to aid recruitment; one of these was an RCT, one a non-randomised prospective comparison and one before and after design. One study attempted to compare two strategies (one passive, one reactive), two compared three strategies and one compared five.

Table 7.5 Summary characteristics of the seven included studies

Factor	Grouping	No of studies
Study design	RCTs	2
	Non-randomised comparisons	2
	Before and after studies	3
Country	USA	7
Disease area	Cancer	3
	Diabetes	2
	Heart disease	1
	Smoking cessation	1
Target groups	Single group	3
	Multiple groups	4
Year of study	Min	1999
	Max	2007
No recruited	Min	33
	Max	8532
No of interventions/strategies	One	3
	Two	1
	Three	2
	Four	0
	Five	1

A more detailed description of the seven selected studies is presented in Table 7.6. Four studies targeted African Americans (Cook, Du, Germino, Harris), two targeted Spanish speaking Hispanics or Mexicans (Kiernan, Martin), with the last targeting all ethnic minorities (Horowitz). Three studies were cancer focussed, two on breast cancer (Du, Germino) and one prostate cancer prevention trial (Cook). All studies were recruiting from the community setting apart from the study reported by Du et al (2009) which recruited from cancer clinics and the study reported by Kiernan et al (2000) which targeted University employees.

Table 7.6 Description of included studies

Author & year	Target population	Design	**Disease area**	Setting	**Country**	Year of study
Cook 2010	African American men vs non- African Americans	Non-randomised prospective comparison	Prostate cancer	Community	USA	2001-2004
Du 2009	African Americans vs Whites	RCT	Breast cancer	Cancer clinic	USA	2003-2005
Germino 2011	African American women vs non-African Americans	Before and after	Breast cancer	Community	USA	Not stated
Harris 2003	African Americans	Before and after	Smokers	Community	USA	1999-2000
Horowitz 2009	Ethnic minority populations	Non-randomised prospective comparison	Pre-diabetes	Community	USA	2007
Kiernan 2000	Hispanic	RCT	Heart disease	University	USA	Not stated
Martin 2011	Mexican	Before and after	Diabetes	Community	USA	Not stated

Table 7.7 provides an overview of the data extracted from all studies. Four of the seven studies included a financial incentive as part of their recruitment strategy (Germino et al, 2011, Harris et al, 2003, Horowitz et al, 2009 and Martin et al, 2011). Four involved or consulted community members to ensure the intervention was culturally appropriate (Cook et al, 2010, Germino et al, 2011, Horowitz et al, 2009 and Martin et al, 2011) and three also addressed language barriers (Germino et al, 2011, Horowitz et al, 2009 and Martin et al, 2011). Neither of the two RCTs used financial incentives, involved community members or addressed language barriers (Du et al, 2009 and Kieran 2000). By contrast, Germino et al (2011), Horowitz et al, (2009) and Martin et al (2011) incorporated all three elements into their strategies.

Four studies reported the number of potential participants screened/approached in addition to the total recruited (Harris, Horowitz, Kiernan and Martin). Only two studies reported rates of retention (Du, Germino). Fours studies presented data comparing the effectiveness of the intervention by ethnic group (Cook, Du, Germino and Horowitz), this was not applicable in three studies where only one group was been targeted (Harris, Kiernan, Martin). All except one study also compared recruitment by strategy (Germino). Finally, only two studies reported data relating to retention (Du et al, 2009 and Germino et al, 2011).

Table 7.7 Overview of the seven included studies

Author & year	Target group	Financial incentive	Community members	Targeted language	Screened /approached	Recruited	Retained	Ethnicity	Strategies
Cook 2010	African American	N	Y	N	N	Y	N	Y	Y
Du 2009	African American	N	N	N	N	Y	Y	Y	Y
Germino 2011	African American	Y	Y	Y	N	Y	Y	Y	N
Harris 2003	African American	Y	N	N	Y	Y	N	n/a	Y
Horowitz 2009	Mix	Y	Y	Y	Y	Y	N	Y	Y
Kiernan 2000	Hispanic	N	N	N	Y	Y	N	n/a	Y
Martin 2011	Mexican	Y[1]	Y	Y	Y	Y	N	n/a	Y

[1] Physicians' offices were paid by researchers to review records and invite potential participants

7.6.4 Characteristics of study populations

One study was targeted at men (Cook et al, 2010), two at women (Du et al, 2009, Germino et al, 2011), and the remaining four were open to both genders. The average age of participants ranged from 43 to 54 years. One study did not provide this information, and two described their target populations as being <50 years or 55 years and greater (Germino et al, 2011 and Cook et al, 2010 respectively).

Three studies reported data relating to socio-economic status, employment or income. Harris et al (2003) found 33% of participants recruited through the proactive phase to be in full time employment compared to 70% of those recruited using reactive methods. This difference was also noted in monthly income with 17% earning a minimum of $1,800 in the proactive arm versus 48% in the reactive arm. Horowitz et al (2009) reported 70% of enrolees to be unemployed, much higher than the already high rate in the general East Harlem population where 52% were unemployed at the time of the research. Furthermore, 45% reported having an annual income of <$15,000 with 49% consequently having no health insurance. The study population can be summarised as middle-aged, Latino, Spanish-only speaking, overweight or obese women. Almost half the participants (48%) in the study reported by Du et al (2009) fell into a low socioeconomic group.

7.6.5 Quality assessment

The quality of the selected articles was found to vary, but were on the whole of intermediate to poor quality. Definitions of trial recruitment reported were found to vary between studies with some reporting numbers enrolled whilst others provided numbers recruited either with or without ineligible participants. Cook et al (2010) reported numbers enrolled but did not explicitly state if all of these patients were then randomised to receive the supplement in this randomised cancer prevention trial (tables 7.8a, 7.8b and 7.8c). In contrast, Du et al (2009) reported proportions willing to enrol in a study before and after exposure to an educational video designed to raise awareness of clinical trials.

7.6.5.1 Study design

Of the seven studies, the five that were not RCTs inherently had potential for selection bias. In addition, neither of the two RCTs provided any information with regards to the method of randomisation utilised (e.g. minimisation, stratified block randomisation), nor was any information given to indicate the intervention allocation list was concealed (Table 7.8). In one RCT, Du et al (2009) randomised patients to view an educational video about research and trial participation or usual care. Kiernan et al (2000) randomised potential participants to one of three mailing strategies inviting them to participate in the study.

All five non-RCTs (two of which were non-randomised prospective comparisons and three before and after studies) were prospective studies and described the patient selection process and inclusion criteria (see tables 7.9 and 7.10). The non-randomised study reported by Cook et al (2010) discussed the effectiveness of recruiting African Americans through the provision of grants to selected sites. Horowitz et al (2009) however, compared the effectiveness of conventional recruitment methods to recruitment by members of the community of interest.

Two of the three studies having a 'before and after' design consisted of a community element (Germino et al, 2011 and Martin et al 2011), see Table 7.10. However, the third study, Harris et al (2003) used ethnicity matched researchers along with incentives for the participants and extended clinic opening hours.

Table 7.8 Quality assessment of RCTs

Bias	Criteria	Author and year	
		Du 2009	Kiernan 2000
Selection bias (RCTs only)	Was the method used to assign patients truly random?	Unclear	Unclear
	Was the allocation to intervention concealed?	Unclear	Unclear
	Comments	Consent taken before randomisation so assume allocation concealed	Randomised to 3 direct mailing strategies which was the initial point of contact
Attrition bias (all studies)	Were at least 80% of patients considered at follow-up?	Yes	Yes
	Was it similar across groups?	Yes	Yes
	Was a valid ITT analysis carried out? (all eligible patients included in the analyses)	Yes	Yes
The intervention (all studies)	Did the design protect against contamination?	Unclear	No
	Did the design protect against performance bias?	Yes	Yes
	Further comments?	Patients in the same waiting room but significant contamination unlikely.	
Relevance (all studies)	Was the nature of the intervention clear?	Yes	Yes
	Was the target of the intervention clearly defined?	Target was all breast cancer patients, there was no tailoring for racial differences but results were analysed by race	Yes
General comments on relevance and applicability		Trial didn't set out to address ethnic minority participation, but does provide information on differences in response to a video by race.	Useful information for trials recruiting Hispanic workers

Table 7.9 Quality assessment of non-randomised prospective comparison studies

Bias	Criteria		Author and year
		Cook 2010	Horowitz 2009
Selection bias	Retrospective or prospective study?	Prospective	Prospective
(uncontrolled/before and	Was the patient selection process described?	Yes	Yes
after studies)	Were details provided of the population from which the sample was selected?	Yes	Yes
	Were there inclusion criteria?	Yes	Yes
	Were all eligible patients invited to participate?	Unclear	Yes
	Is it possible that the investigators had discretion over who was selected?	Unclear	Yes
Attrition bias (all studies)	Were at least 80% of patients considered at follow-up?	Unclear	Unclear
	Was it similar across groups?	Unclear	n/a
	Was a valid ITT analysis carried out? (all eligible patients included in the analyses)	Unclear	n/a
The intervention (all	Did the design protect against contamination?	No	Unclear
studies)	Did the design protect against performance bias?	No	Unclear
	Further comments?	No protection against biases built in	Will participants respond more favourably to members of their community vs conventional methods
Relevance (all studies)	Was the nature of the intervention clear?	Yes	Yes
	Was the target of the intervention clearly defined?	Yes	Yes
General comments on relevance and applicability		Good comparative data between grant awarded and non-awarded sites as well as pre and post grant data	Relevant paper concerning CBPR methods and how they could be used with great effective for hard-to -reach populations. Unfortunately not a randomised comparison and small numbers but good basis for further work

Table 7.10 Quality assessment of before and after comparison studies

		Article author and year of publication		
Bias	Criteria	Germino 2011	Harris 2003	Martin 2011
Selection bias	Retrospective or prospective study?	Prospective	Prospective	Prospective
(uncontrolled/ before and	Was the patient selection process described?	Yes	Yes	Yes
after studies)	Were details provided of the population from which the sample was selected?	No	Yes	Yes
	Were there inclusion criteria?	Yes	Yes	Yes
	Were all eligible patients invited to participate?	N/a	Yes	Yes
	Is it possible that the investigators had discretion over who was selected?	Yes	Yes	Yes
Attrition bias (all studies)	Were at least 80% of patients considered at follow-up?	Yes	Unclear	Unclear
	Was it similar across groups?	Yes	N/a	Unclear
	Was a valid ITT analysis carried out? (all eligible patients included in the analyses)	Unclear	N/a	Unclear
The intervention (all studies)	Did the design protect against contamination?	Unclear	No	Unclear
	Did the design protect against performance bias?	Unclear	N/a	Unclear
	Further comments?	Partial CBPR approach partnering with community leaders /organisations	Thought out design e.g. ethnicity matched researchers, incentives to attract low income population, evening appointments etc	CHWs useful in recruitment as they have trust within the community
Relevance (all studies)	Was the nature of the intervention clear?	Yes	Yes	Yes
	Was the target of the intervention clearly defined?	Yes	Yes	Yes
General comments on relevance and applicability		Not known which parts of the strategy were the most effective	Good comparison data although not clean comparison e.g. the reactive phase covered a wider area than the proactive phase	Good before and after data showing the efficiency of CHWs and their role in recruitment

7.6.5.2 Selection and attrition bias

Selection bias occurs when there are systematic differences between the selection of participants to different treatment groups. Attrition bias refers to systematic differences in withdrawal rates, or exclusion from either the study sample or analysis across treatment groups. It is not clear from the information provided whether the RCTs were subject to selection bias (Table 7.8). However, for the five non-RCTs, it is possible that researchers may have inadvertently selected people who were either more or less likely to participate, thus influencing the results which could lead to either an under or overestimation of the effectiveness of the intervention (Table 7.9 and 7.10). Four of the five non-RCTs described the population from which the participants were selected. It was not clear whether all eligible patients were asked to participate in two of five non-RCTs.

7.6.5.3 Performance bias

Given the nature of the interventions, blinding was not feasible in any of the seven studies which may have resulted in an element of performance bias. Performance bias can be defined as systematic differences in the way in which participants are treated or interacted with, other than the intervention of interest. In this particular field, this could have been recruiters spending more time with potential participants, providing fuller answers or generally having a friendlier approach. Performance bias is minimised by ensuring interventions are delivered in a standardised way by all recruiters or researchers. The two RCTs (Du et al, 2009 and

Kiernan et al, 2000) were protected against performance bias through their study design, but whether this had been done was unclear or not known in the remaining five studies (Table 7.9 and 7.10).

7.6.5.4 Contamination

Contamination between groups is another important problem which can occur when the same researcher is delivering more than one intervention and an element of one intervention leaks into the other. It can also occur through contact between potential participants. No steps to minimise contamination were reported in any of the seven studies and so it is possible that the effects reported may have been exposed to this bias (Table 7.8, 7.9 and 7.10).

7.6.5.5 Summary of the quality of the selected studies

To summarise, the selected studies were of intermediate to poor quality with little evidence to suggest bias had been taken into account. The studies were varied in design with only two identifying themselves as RCTs, although, no information regarding the randomisation method used was provided. Given the differences in the types of study and the small number of studies meeting the selection criteria it was not possible to perform any further analysis, such as meta-analysis, using these data. The reliability of these data is questionable given the large potential for bias.

In three of the seven included studies, a minimum of 80% of patients were still being followed up in the study at time of reporting (Du et al, 2009; Kiernan et al, 2000 and Germino et al, 2011). All three studies also reported the percentage follow up to be similar across groups. None of the seven articles described any measures to minimise contamination. One study protected against performance bias by delivering a standardised intervention in the form of an educational video (Du et al, 2009). The intervention itself and the group the intervention was aimed at were described in all seven studies (tables 7.8, 7.9 and 7.10).

7.7 Synthesis

The interventions tested were varied but can be grouped into five themes:

Community outreach

- Use of cultural brokers
- Partner-led recruitment
- Community based/recruitment events
- Recruitment at targeted local organisations
- Faith based promotion

Advertising

- Flyers
- Adverts in targeted outlets/centres
- Media (radio/television/newspaper adverts)

Incentives

- Government grants
- Financial incentives for staff and participants
- Non-financial incentives

Targeted mailing

- Direct mailings (e.g. postcodes with high African American population)

Referral

- Mailshots through health insurers
- Referral by physician

Further information relating to the interventions used can be seen in Table 7.11 and the effect upon recruitment by ethnic group and strategy in Table 7.12. As interventions were used in various combinations and given the small number of studies each is described in detail in order to provide a clearer picture.

Table 7.11: Interventions used in included studies

Author & year	Intervention/strategy	Financial incentive	Community members	Targeted language
Cook 2010	Grant to boost recruitment. $1.1m in total awarded to 15 sites by the National Cancer Institute	No but funds were used to aid parking and transportation and provide refreshments at recruitment meetings	Sites awarded the funding consulted community leaders and spokes people but reported these as less successful strategies	No, targeted AA
Du 2009	Educational video	No	No	N/a (AA/Whites)
Germino 2011	Increasing familiarity with study, and availability and accessibility of study literature through cultural brokers	$20 at each data collection plus a small gift every 2 months	Yes. Worked with community organisation who trained lay health advocates. Also used cultural brokers and involved pastors	Language and tone in materials adjusted for young AA audience
Harris 2003	Personal approaches (proactive) vs flyers, media, church based promotions (reactive)	$100 across 3 visits ($20,40,40) plus small promotional gifts e.g. t-shirts, tote bag, key chain, water bottle, pen, magnet. Participants were also presented with certificates of accomplishment.	No but community-based research at well know community health centres using AA researchers	N/a
Horowitz 2009	1. Physician referral, 2. recruiting at large events, 3. holding recruitment events, 4. recruitment at local organisations, 5. partner-led recruitment	$50 gift certificates. Small gifts such as study t-shirts and pens were also offered as well as a 'healthy box lunch'	Community-based participatory research (CBPR), collaboration between community and academia/researchers. Developed and tested 5 community led strategies.	English and Spanish
Kiernan 2000	1. Flyer alone, 2. Flyer plus personalised letter, 3 Personalised letter specifically mentioning Hispanic heart disease risk	No	No	No
Martin 2011	1. Community outreach, 2. adverts, 3. recruiting through insurers	Financial incentive offered to Physician offices to review their records and invite any potential participants	Yes. Community Health Workers (lay people from the target population) hired to deliver the diabetes self-management intervention in this trial joined the researchers to aid recruitment.	Bilingual ads and personnel appearances

Table 7.12: Recruitment by retention for included studies (continued on the following page)

Author & year	Approached/ screened	Recruited	Retained	Recruited by ethnicity	Recruited by strategy	**Before and after comparison**
Cook 2010	-	8532	-	Non AA=6355 (76%) AA=2177 (26%) Total=8532	All mean participants per month at grant sites went from 8.8 to 9.1 compared to 6.8 to 3.0. Increase observed in grant/non grant sites in AA (2.4 to 2.9 vs 1.8 to 1.1)	Mean accrual of AA participants increased from 34% pre to 44% post grant. In comparison sites the increase for the same period was 25% to 28%. All mean participants per month at grant sites went from 8.8 to 9.1 compared to 6.8 to 3.0. An increase was also observed between grant/non grant sites in AA (2.4 to 2.9 vs 1.8 to 1.1)
Du 2009	-	218	98% (193)	AA=89 (45%) vs Whites=107 (55%)	At baseline 23% whites vs 12% AA indicated that they would be extremely likely to enroll. At FU this was 22% vs 15%.	N/a
Germino 2011	-	104	87%	AA made 31% of the final sample (over-represented by 10%. State AA pop=21%)	No comparison by strategy. Overall figures presented.	Increased from 22 pre-strategy to 104 post strategy
Harris 2003	1490	600	-	None as AA only	Proactive phase resulted in 66 pts, reactive in 534 pts (18% and 48% of those screened). Newspaper and TV ads gave the best yields (59% and 58%), the Health Centre gave the worst at 20.3% (% from no's screened)	Proactive phase=66 recruits Reactive phase=534 recruits

Horowitz 2009	555	99	-	89% Latino, 9% black, 1% Asian, and 1% Native American	99 enrolled, 18% of approached. Clinician referral=0, Special events=17%, Community organisations=9%, Public events=6%, Partner-led=34%	n/a
Kiernan 2000	561	38	-	n/a as all Hispanic	Flyer alone=2.1%. Flyer + general population risk statistics for heart disease=6.5% Flyer + Hispanic population risk statistics for heart disease=9.1%	n/a
Martin 2011	343	144	-	n/a, all Mexicans	53.5% of all patients randomised came from CHWs. Church/community events were least successful at 3.5%. When looking at proportions recruited by source family/friend (word of mouth) was the most efficient with 55% of folks screened being randomised. Followed closely by CHWs at 53%. Community events were least efficient at 23%.	Twelve months were allocated to complete recruitment of 144 participants. The insurer enrolled 259 Hispanics with diabetes, but despite vigorous attempts (up to 10 letters/phone calls) only one participant was randomized.

Du et al (2009) who focussed on the willingness of breast cancer patients to enter a trial found attitudes did not change significantly post video intervention. Kiernan *et al* (2000) reported improvement in recruitment rates in her RCT. Hispanic University employees were mailed either 1) a flyer about heart disease, 2) a flyer containing heart disease risk statistics for the general population, or 3) a flyer containing heart disease risk statistics for the Hispanic population. Results show 2.1% were recruited through method 1, 6.5% via method 2 and 9.1% via method 3. No incentives were offered in this study. Although, performance bias was controlled for by the design of the study it is unclear whether there was an element of contamination.

Cook et al (2010) reported that government grants to enhance the recruitment of ethnic minorities improved recruitment into their study. Comparisons were made pre and post award for grant sites as well as between grant awarded and non-grant awarded sites. Reimbursement of travel and parking fees were available and refreshments were provided free of charge at recruitment meetings. Cook reported mean accrual of all participants to increase from 8.8 to 9.1 participants per month pre and post grant award compared to 6.8 to 3.0 in non-awarded sites in the same time window. The change was also apparent when examining African American groups alone (2.4 to 2.9 in awarded sites vs 1.8 to 1.1 in non-awarded sites).

Martin et al (2011) offered physician's offices a financial incentive to review their records and invite any potential participants. This method was unsuccessful, resulting in the randomisation of one participant out of a possible 259. In contrast, enlisting the help of local Community Health Workers yielded 53.5% of all recruited

participants. Horowitz et al (2009) conducted a Community-Based Participatory Research (CBPR) exercise in which researchers and community members developed five recruitment strategies. Participants were also offered $50 gift certificates, small promotional gifts and a healthy packed lunch. Recruitment through partner-led organisations was the most successful yielding 34% of participants. Clinician referral was the poorest strategy resulting in no enrolees.

Proactive methods rely upon recruiters seeking out and inviting potential participants whereas reactive recruitment methods rely upon the potential participants to make contact with the research team if they are interested in participating. Harris et al (2003) set out to compare the strategies using a before and after design with the first stage being proactive recruitment from a Health Centre. This resulted in 66 recruits (18%) compared to the latter phase where 534 patients were recruited (48%) following a newspaper and television advertising campaign. The comparison is not a clean one as both strategies were in operation in the latter phase coupled with the fact that the proactive phase was restricted to the health centre whereas the methods used in the reactive phase reached a much larger population e.g. television advertisements and large mailshots. This suggests strategies that enable researchers to reach larger populations may be more efficient in attracting potential participants whereas individual orientated approaches may be more effective once a patient has expressed an interest. Participants were also offered a total of $100 payment as well as promotional gifts

which may have improved uptake. Certificates of achievement upon completion were awarded, but rates of retention were not reported in this article.

Germino et al (2011) used cultural brokers, worked with community organisations that trained lay health advocates and enlisted the help of pastors. In addition, the language and tone of the study materials were enhanced to be more appealing to this young African American breast cancer population. Supplemented with a financial incentive at each data collection point plus small gifts every couple of months resulted in 22 recruits pre-interventions and 104 post. However, the article did not break recruitment down by strategy so it is not possible to decipher if a particular element was more effective. Based upon pilot study data Germino et al (2011) reported African American women to be more likely to be single parents than their Caucasian counterparts and hence the researchers put great effort into working around the women's availability. Furthermore, participants were sent birthday cards and gifts letting them know how valuable their contribution was in an effort to boost retention. By the end of the study 87% of participants had been retained.

Two of the seven selected trials addressed non-English speaking ethnic minorities whilst four focussed on African Americans. The single most significant barrier to African American participation is well documented to be mistrust, a legacy left by the Tuskegee experiment (Brawley, 1998). As demonstrated by Harris et al (2003) and Martin et al (2011) using ethnicity matched researchers appears to be an

effective three pronged tool as it not only bridges language and cultural gaps but also trust issues.

Cook et al (2010) demonstrated how grants may be a good way of improving the recruitment of ethnic minority groups because it gives the site the flexibility to use the grant in ways which are appropriate for their local population. The grants were not awarded randomly. Instead a points system was used which took into consideration the proportions of ethnic minorities served by the site and any track record of recruiting these populations into research. The authors emphasised the need to have grants and any interventions in place prior to the start of recruitment.

Trust issues, particularly within the African American communities of the US dominated the literature (Cook et al, 2010; Du et al, 2009; Germino et al, 2011; Harris et al, 2003). The Tuskegee study of syphilis conducted between 1932 and 1972 which left African American males untreated is considered to be the main source of distrust and suspicion of healthcare professionals in the US (Brawley, 1998).

The motivation to include ethnic minorities in the USA was driven by a change in legislation. The NIH Revitalization Act (1993) which stipulates all minorities including women, who were routinely excluded from research in the aftermath of the thalidomide catastrophe, to be represented in clinical trials in accordance with the representation of these groups in a geographical area or proportion suffering with the condition of interest. To ensure adherence to this policy the NIH linked

recruitment to funding provision with trials failing to recruit representative samples of women and ethnic minority groups having their funding terminated (NIH, 2001 and Air Alert, 1998).

The motivation to participate in clinical trials in the USA may differ from the UK where access to health care is freely available on the NHS. Patients without insurance or with inadequate cover in the USA may be more likely to participate in a clinical trial in order to benefit from free health care provision. Further to this, the USA also has a different attitude to the use of financial incentives to attract trial participants compared to the UK where payment above and beyond expenses occurred as a result of participating in a trial are classed as inducement. Three of the seven selected studies offered financial incentives to the participants (Germino et al, 2011; Harris et al, 2003; Horowitz et al, 2009), with a further study offering a financial incentive to healthcare professionals who invited patients to participate (Martin et al, 2011).

To summarise, interventions that were reported to be effective in facilitating the recruitment of ethnic minority participants included the provision of government grants to this end. Community health workers and ethnicity matched researchers were reported to improve recruitment rates along with targeted mailshots, and television, radio and newspaper campaigns. As most studies used a combination of intervention or strategies it was not possible to obtain a clear picture of the effectiveness of each method in turn. Authors reported recruiting through health

insurance companies, at church events and health fairs to be the least successful strategies.

Further to the article by Kiernan et al (2000), results from a study of two direct mailing strategies in a weight management trial have recently been published by the same group (Brown et al, 2012). The articles reports potential participants to be more likely to respond to a letter inviting them to participate in a study containing an ethnically targeted statement compared to a personalised letter.

7.7.1 Limitations of the review

Extensive searches were performed to identify potential articles for this review. However, the lack of indexed terms for ethnic minorities resulted in the search strategies relying heavily upon keyword searching which limited it to the common terms used by the authors. As a consequence it is possible that studies may have been missed. The review was also restricted to articles written in English and so it is possible that relevant studies published in other languages have been missed.

Further to this, this review focused on interventions to improve recruitment and retention of ethnic minority and under-represented groups into clinical trials. Screening trials were excluded as it was felt that motivations for entering a screening trial would be different to that of other trials, e.g. disease specific.

Lastly, the quality assessment criteria for attrition bias were included in order to ascertain the quality of the studies and therefore the reliability of the evidence on the effectiveness of the reported interventions. However, as the literature review was focussed upon interventions that improved both recruitment and retention the quality checklist for attrition bias was not the most appropriate tool and alternatives would need to be found.

7.8 Conclusion

Very few studies met the inclusion criteria and only two of these were RCTs. The interventions to improve recruitment and retention were diverse, which is expected given the complexity of barriers to participation which can differ by a combination of many factors, e.g. ethnic group, age, disease, country, healthcare system, history/past experiences. On the whole, the systematic review highlighted the lack of research in this area and emphasised the need for good quality studies in the future.

All the included studies originated from the USA and, although they are of interest, not all interventions carried out would be feasible in the UK, e.g. the use of financial incentives in patients in particular. Payment is limited to subsistence cost, such as travel, parking fees, loss of earnings, carers cost if applicable or other costs that may preclude people from participating in research. Paying a patient in excess of this could be deemed coercion and would need to be approved by an ethics committee. Payment may affect the sampling process as patients who are

struggling financially may be more willing to participate than if the payment was not in place. This could therefore jeopardise the generalisability of the study. Financial incentives would also be an issue in other countries outside of the USA, such as Australia where, for similar reasons to the UK, it would violate ethical principles and be considered inducement (National Statement of Ethical Conduct in Human Research, 2007).

Over half of the selected literature was dominated by building trust in African American communities; the effects of Tuskegee may not be as apparent in the UK. Although there was some evidence of interventions having a positive effect on recruitment, the evidence was not of sufficient quantity or quality to come to any definitive conclusions. However, this review has provided a basis for further work.

Identifying effective strategies to facilitate the recruitment and retention of ethnic minorities in UK trials is needed given the ageing population structure of these groups. In the future, we expect to see more cancers as well as other diseases, just as we have witnessed the growing rates of diabetes and heart disease in ethnic minority populations. Kiernan et al (2000) states, “if suitable interventions are not identified, researchers may fail to recruit representative populations causing delays not only in recruitment time but also in progress into minority health”. The effectiveness of interventions may differ by ethnic minority group but I am unable to comment on this topic given the thin evidence obtained through this systematic review.

Based upon findings from the literature review above I would recommend future researchers wishing to trial interventions to improve the recruitment and retention of ethnic minority groups into RCTS in the UK to consider a cluster design with sites trialling different strategies coupled with a before and after comparison as demonstrated in the study by Cook et al (2010). Sites would be matched in terms of proportions of ethnic minority populations served by the site and the incidence of disease within ethnic minority groups e.g. diabetes more prevalent in South Asians compared to the Chinese population.

Interventions I would consider for the UK, where payment beyond expenses incurred as a direct result of trial participation would be not considered acceptable, would include targeted mailshots in appropriate languages, adverts in local businesses and local radio stations. Ethnicity matched researchers, link workers or patient navigators who have knowledge of the study and/or disease area and can guide a potential participant through trial concepts, e.g. equipoise and randomisation, the study protocol and answer any questions. Early consultations and input from representatives of ethnic minority community groups when devising strategies and securing funding in order to ensure sufficient resources, e.g. the link workers would be advised.

Chapter 8 Under-representation of ethnic minorities in cancer clinical trials: assessing the magnitude of the problem

8.1 Introduction

Randomised clinical trials are improving treatment in many disease areas, cancer being one of them. There are many benefits for patients who participate in an RCT which include access to new drugs or interventions before they are available on the NHS. There is also evidence to suggest that patients who participate in clinical trials have better outcomes (Braunholtz et al, 2001). This may be due to closer monitoring of their disease and general wellbeing as part of the trial protocol. However, the efficacy of treatments and toxicities suffered may vary between ethnic groups due to differences in pharmacokinetics (what the body does to the drug) and pharmacodynamics (what the drug does to the body). Ethnic factors can differ in two ways, intrinsically or extrinsically. Intrinsic factors include comorbidities and genetics, for example, the discovery of the salt sensitivity gene in African Americans (Johnson 1997, Xie et al, 2001, McGraw and Waller 2012). Extrinsic influences include environmentally factors such as diet, exercise and smoking habits.

Ethnic minorities have been shown to be poorly represented in clinical trials (Hussain-Gambles et al, 2003 and 2004, Mason et al, 2003, Murthy et al 2004, Godden et al, 2010), although there has been little recent evidence from the UK to support this. More recent research from the USA has been published which is of interest but not completely relevant to the UK (Hoel et al, 2009, Geller et al, 2011). The USA differs in several aspects:

1) The proportion of ethnic minorities is much higher, in 2010 22% identified themselves as belonging to a ethnic minority group (US Census bureau, 2010)

2) The ethnic minority groups are different (the USA has a large proportion of Hispanic, Pacific Islanders and Alaskan Natives as well as African and Asian Americans)

3) The healthcare system is privatised which makes receiving treatment difficult or impossible if you do not have adequate health insurance

The importance of collecting ethnicity data has been increasingly recognised since the registration of this PhD. Cancer Research UK highlighted the importance of collecting ethnicity data in cancer patients and in 2006 commissioned a project to identify methods by which ethnicity data collection could be improved for cancer statistics.

Cancer registration data is collected by 11 registries in the UK. The data is obtained from a range of sources, including hospitals, cancer centres, pathology laboratories,

private hospitals, general practice, hospices, nursing homes, screening programmes and death certificates. Data collected included tumour and patient characteristics but ethnicity was not included in the dataset until a few years ago.

Data linkage was one possible means of obtaining ethnicity from Hospital Episode Statistics (HES) datasets that has been explored by cancer registries. The HES database holds information on every hospital encounter a patient has as either in- or outpatient. Ethnicity has been collected as part of HES data since 2004/5 and although collection has been patchy there have been improvements over recent years (HES online, 2004 and 2011). One of the main issues encountered during the linkage exercise was the occurrence of conflicting ethnicity codes where patients had multiple records. Methods using the 'most frequently recorded ethnicity' or the 'last recorded ethnicity' were explored as well as multiple imputation to overcome this (Ryan et al, 2012).

In 2009, the National Cancer Intelligence Network (NCIN) in collaboration with cancer registries published a report of cancer incidence figures using this newly collated data (NCIN, 2009). The report is the first of its kind, but is based upon incomplete data. Ethnicity was reported to be missing/unobtainable for approximately 25% of cancer registrations and the findings were limited to broad ethnic categories: White, Asian, Black, Chinese, Mixed and Other due to the small number of cancer registrations of patients in the ethnic minority groups. Three levels of sensitivity analyses were performed as a consequence of the missing ethnicity data, the first assuming patients with missing ethnicity were missing

completely at random therefore had the same distribution as the patients with recorded ethnicity. This was followed by extreme scenarios, firstly assuming all the unknowns were of White origin which produced the lowest rates of non-White incidence and secondly assuming they were non-White giving the highest rates. The results proved tricky to interpret.

The aim of this chapter was to attempt to assess the size of the under-representation problem by estimating how many ethnic minority patients would be expected to go into cancer trials using data from the following sources which were available for use in this thesis:

1. 2001 and 2011 UK census
2. Hospital Episode Statistics audit data from a large local teaching hospital
3. Cancer Research Network trial recruitment data from one ethnic rich hospital

These datasets were chosen to obtain rates of the ethnic minority populations in the UK, England and Wales and the local West Midlands area, and to examine ethnicity data collected in one local hospital and also to assess rates of cancer by ethnic group in this ethnic rich population. The local cancer research network recruitment dataset (also from the same ethnic-rich population) was chosen to enable assessment of the representation of ethnic minorities in cancer clinical trials.

8.2 Methods

8.2.1 Census data

8.2.1.1 Census 2001

Population pyramids of the 2001 UK inhabitants exploring age, gender and country of birth across ethnic group were produced using Microsoft Excel to study the changing make-up of the UK population and the West Midlands population independently. The data presented here is from a table commissioned from the ONS in 2009. It included data from the Scottish and Northern Ireland censuses for the same period. The ethnic categories used by each census differed slightly (as explained in Chapter Two) and were harmonised to form 11 common groups for this analysis.

8.2.1.2 Census 2011

Data from the recent release of 2011 census was extracted from the ONS data table KS201EW which provides information split by local authority and also split by ethnicity. These data were used to compare the distribution of Coventry's ethnic population to that of England and Wales. Only limited information was available for the 2011 census at the time of completing this thesis. As ethnicity was not broken down by age in census table KS201EW it was not possible to calculate age-specific incidence rates.

8.2.2 Hospital Episode Statistics

Audit datasets for the period Jan 2007 to Dec 2012 were made available for use in this thesis. Completeness of ethnicity fields in both inpatient and outpatient datasets was assessed. Place of Birth, Religion and Language data were also compared across datasets.

Records meeting the following criteria were selected to create four datasets:

8.2.2.1 Inpatients (dataset A)

Patients with International Classification Disease codes for NEOPLASMS (ICD version 10, codes C00 to D49) were identified and selected. Inpatient data were read into SAS statistical software. Patients had anything between 1 and 16 diagnoses, with many of these being non-cancer. However, some cancer codes referred to subsequent malignancies or disease metastasising to other areas of the body coded as 'cancer of unknown or ill-defined origin'.

These data were examined in the first instance looking purely at primary diagnosis (one row per patient format) followed by all diagnoses (multiple rows per patient). Non-cancer codes were re-coded to missing using SAS statistical software. I transposed this dataset in SAS in order to obtain the frequency of cancers suffered per patient. Cancer diagnoses were explored using frequency tables and cross-tabulations. Scrutinizing all diagnoses provided a clearer picture of the prevalence rates of disease by site.

Incidence is a commonly used measure of the risk of developing a disease within a specified time frame. Precise incidence rates are very difficult to estimate even when not considering ethnic group. For this reason I have reported rates of cancer over a six year period (2007 to 2012) and one year period (Jan to Dec 2012).

8.2.2.2 Outpatients (dataset B)

The outpatient records originated from all oncology clinics held at a large local teaching hospital between 2007 and 2012. Data were provided in a one row per patient format with duplicates removed where possible. Further details relating to the clinic were not available.

As it was possible for patients to be treated exclusively as inpatients, outpatients or indeed both then it was necessary to inspect these data in two additional permutations (datasets C and D below).

8.2.2.3 Overlapping set (dataset C)

Patients treated as both inpatient and outpatients presented an opportunity to compare reported ethnicities. The data were read into SAS statistical software and merged using a unique identifier. Rates of agreement for ethnic group across datasets were reported.

8.2.2.4 Final set (dataset D)

Inpatient records were merged in SAS with the records of patients treated as outpatients only to form the final dataset of all records. For patients with missing/unknown ethnic group the record with most information was taken. This dataset was used to calculate the cancer rates for primary diagnoses and transposed using SAS to obtain all diagnoses.

Exploration of the inpatient and outpatient data sets revealed two separate codes for ethnic origin, local and national (which is in keeping with the 2001 census categories). The codes were identical except for the inclusion of 'not known' and 'not specified' in the local coding which is not offered in the national list. The local ethnicity coding was used for the entirety of this research.

I examined these data over the full six year period and for 2012 exclusively to allow comparison with 2011 census data for Coventry and recruitment data from the selected Cancer Research Network.

8.2.3 Cancer Research Network recruitment data

Locally a Cancer Research Network (CRN) entered 35% of eligible patients into trials in 2005/6 (NCRN, 2006). In April 2012, the network changed software to a new database (EDGE) which gave them the ability to record information on all potential participants seen by the research team and their trial status (recruited, screened, declined). The database also had dedicated fields to record ethnicity and reason for

non-participation. Ethnicity was transferred from the HES database or medical notes to the recruitment database without consulting the patient. These data were available for April 2012 to April 2013 at the time of this write-up. Trial status was explored by ethnicity to gauge the numbers of patients invited to participate, those declining and numbers recruited.

This dataset in conjunction with the final dataset derived from HES (dataset D) were used to crudely assess representiveness of cancer patients in clinical trials in the local area.

8.2.4 Assessing representation

In order to estimate the representativeness of the patients coming through the doors of the selected Cancer Research Network based within a large local teaching hospital, the total numbers of patients attending the hospital with cancer as reported above were used to form a denominator. The Cancer Research Network recruitment data were only available from April 2012 to April 2013 whereas HES data were available from Jan 2007 to Dec 2012. In order to assess representativeness over a year I chose to use HES data for 2012 (Jan-Dec) as it maximised crossover with the CRN recruitment data.

In order to obtain a more accurate denominator I deleted patients meeting any of the following three criteria from the 2012 HES data. Firstly, patients dying in hospital as they may not have lived long enough to be identified by the research network, invited to participate in a trial or may have been filtered out at multi-

disciplinary team (MDT) meetings because they were not fit enough to tolerate the treatment or did not meet other eligibility criteria. A patients' treatment package is usually decided at these meetings and any potential trial candidates identified. The MDT may take a variety of criteria into consideration before making their decision, for example, co-morbidities or the ability to give informed consent. A list of open trials and their inclusion and exclusion criteria are usually available at these meetings which are frequently attended by a member of the research team.

Secondly, patients who discharge themselves may not be interested in receiving treatment and so it could be assumed that there is a higher probability that they may not be in hospital long enough to be invited to participate in a trial, or would not participate if invited. This could also be true of those discharged by a relative. For this reason, these patients did not form part of the denominator. Thirdly, it was hypothesised some patients admitted for surgery may not be identified as possible candidates, e.g. cancer removed with no further treatment. These patients were also excluded from the denominator.

8.3 Results

8.3.1 City X's ethnic population

The population of City X is very diverse, with a higher proportion of non-white minorities than reflected in the overall census distribution (26.2% compared to 14.1% respectively, see Table 8.1). This does not take into account the recent influx of Polish natives who would be placed within the 'Other White' category, which is

higher in City X than nationally (4.9% vs 4.4%). The largest non-white ethnic minority groups in 2011 England and Wales were Indian, Pakistani and Black African accounting for 2.5%, 2% and 1.8% respectively. Although the City X top three ethnic minority groups are the same as England and Wales, the proportions are different with Indians making up 8.8%, Black Africans now in second place with 4% and Pakistanis' accounting for 3% of the population.

Table 8.1 England and Wales and City X, 2011 census by ethnic group

		England & Wales		City X	
Category	Ethnic group	N	%	N	%
White: British, Scottish,	White British	45,134,686	80.5	211,188	66.6
Welsh, Northern Irish	White Irish	531,087	0.9	7,305	2.3
	White Gypsy or Irish traveller	57,680	0.1	151	0.0
	Other White	2,485,942	4.4	15,385	4.9
Mixed/Multiple	White and Black Caribbean	426,715	0.8	3,672	1.2
Ethnicities	White and Black African	165,974	0.3	943	0.3
	White and Asian	341,727	0.6	2,388	0.8
	Other Mixed	289,984	0.5	1,227	0.4
Asian/Asian British	Indian	1,412,958	2.5	27,751	8.8
	Pakistani	1,124,511	2	9,510	3.0
	Bangladeshi	447,201	0.8	2,951	0.9
	Chinese	393,141	0.7	3,728	1.2
	Other Asian	835,720	1.5	7,658	2.4
Black/African/	African	989,628	1.8	12,836	4.0
Caribbean/Black British	Caribbean	594,825	1.1	3,317	1.0
	Other Black	280,437	0.5	1,611	0.5
Other ethnic group	Arab	230,600	0.4	2,020	0.6
	Any other ethnic group	333,096	0.6	3,319	1.0
Total	All categories	56,075,912	100	316,960	100
	Non-White	7,866,517	14.1	82,931	26.2

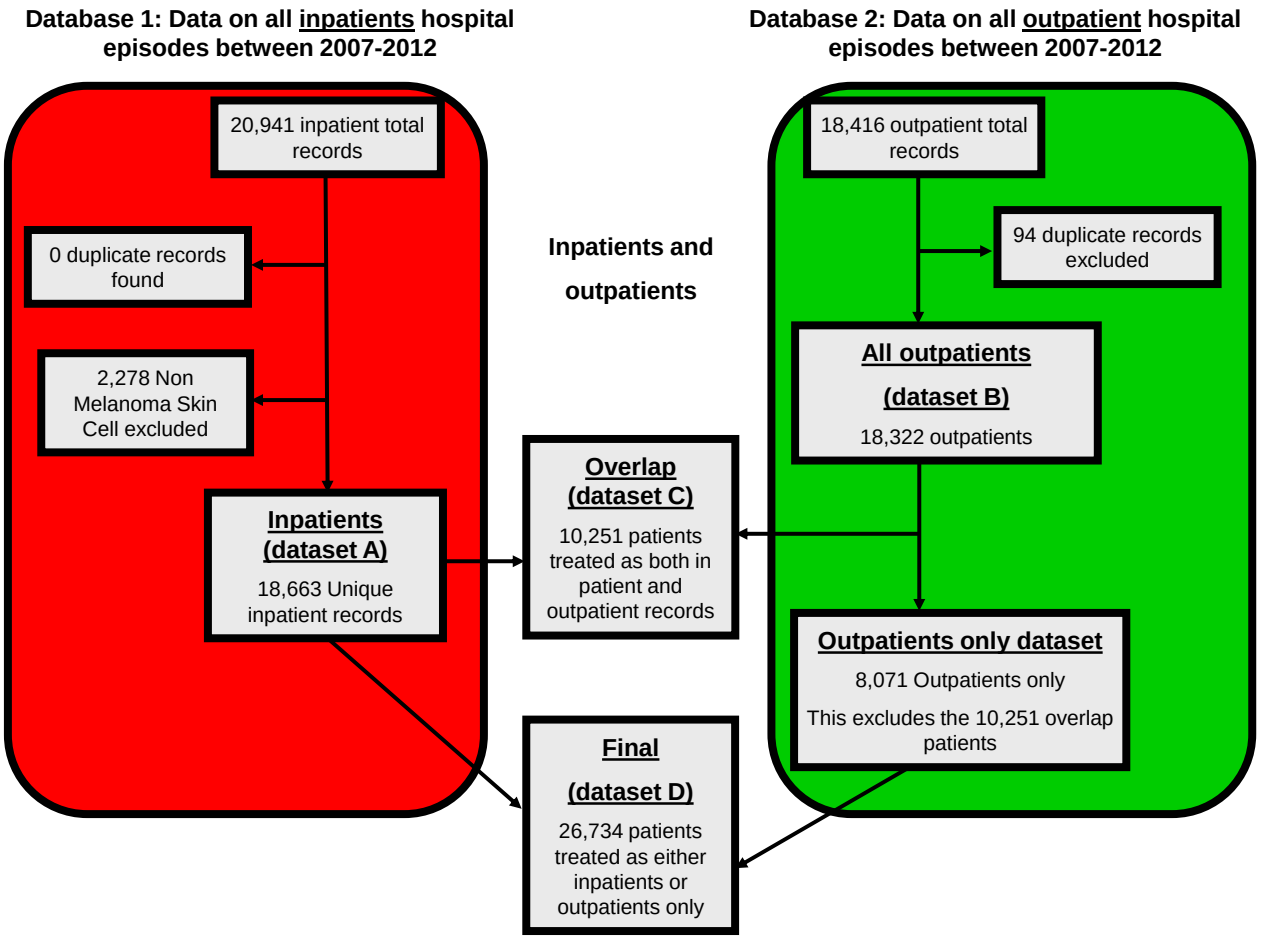
Note: Data extracted from ONS 2011 census data table KS201EW.
(http://www.ons.gov.uk/ons/search/index.html?newquery=ethnic+group)

8.3.2 Hospital Episode Statistics

Analysis of ICD codes revealed 2278 patients with Non-Melanoma Skin Cancer (NMSC) and no further cancers. These patients were excluded from analysis as they are easily treated and are very common (usually abnormal moles and similar). They are slow growing and prognosis is very good hence the dearth of research in this area. This left 18,663 records remaining in the final inpatient dataset (dataset A, see Figure 8.1).

Ninety four duplicate records were identified in the outpatient dataset and excluded leaving 18,322 records in the final outpatient dataset (dataset B). In terms of combined datasets, 10,251 patients were found to have a record in both in and outpatient sets (dataset C). The final dataset comprised of all inpatient records merged with patients who were treated as outpatients only (18663 + 8071=26,734, dataset D).

Figure 8.1 Process of HES dataset creation



8.3.3 Completeness of ethnicity in inpatients and outpatients

The inpatient population (dataset A)

The majority of inpatients were White British as expected making up 86.08% of the population (Table 8.2). The second largest group, although considerably smaller, were Indian (2.75%). Alarmingly, the third largest category were 'not known' comprising 2.44%, although 'not specified' and 'not stated' codes contributed a further 0.03% and 1.26%, respectively, resulting in total unknown ethnicity codes of 3.73%. The Irish group contributed 2.34% whilst the Other White group was the next largest at 1.65% which included residents of Eastern European origin (e.g. Polish). The smallest groups were those who chose not to specify their ethnicity (0.03%), Chinese (0.14%), Bangladeshi (0.16%), Other Black (0.22%), Other Asian (0.33%), African (0.35%) and Pakistani (0.5%).

The outpatient population (dataset B)

There appeared to be a smaller proportion of White British outpatients compared to inpatients (78.17% vs 86.08%) see Table 8.2. Furthermore, All Whites accounted for 81.61% of outpatients yet 90.07% of inpatients. This anomaly may simply be due to the poorer collection of outpatient data. The outpatient ethnicity data contained a greater proportion of missing and not known, not specified and not stated codes accounting for 13.21% of outpatients compared to 3.73% for inpatients. The proportions of the remaining non-White ethnic minorities were similar across both datasets.

Table 8.2 : Datasets A and B-Ethnicity coding of the inpatient and all outpatient records

	A=Inpatients		B=Outpatients	
	N	%	N	%
A=White British	16065	86.08	14322	78.17
B=Irish	436	2.34	378	2.06
C=Other White	308	1.65	253	1.38
D=Mix White+ Caribbean	22	0.12	27	0.15
E=Mix White + African	15	0.08	16	0.09
F=Mix White + Asian	4	0.02	8	0.04
G=Mix Other	15	0.08	10	0.05
H=Indian	514	2.75	437	2.39
J=Pakistani	93	0.50	60	0.33
K=Bangladeshi	30	0.16	18	0.10
L=Other Asian	61	0.33	52	0.28
M=Caribbean	131	0.70	106	0.58
N=African	65	0.35	43	0.23
P=Other Black	41	0.22	37	0.20
R=Chinese	26	0.14	26	0.14
S=Other ethnic	140	0.75	107	0.58
NKN=Not known	456	2.44	1430	7.80
NSP=Not specified	6	0.03	42	0.23
Z=Not stated	235	1.26	949	5.18
Total	18663	100	18322	100

The split between males and females was roughly equal in the inpatient population with 49% being female (see Table 8.3). However, an imbalance was noted in the outpatient set, with 57% of patients being female. A large proportion of the patients were over 60 years of age (70% of inpatients vs 64% of outpatients) which was not unexpected in this cancer population. These data show a greater proportion of females than males being treated at a younger age in both sets (inpatients=35% females and 24% males, outpatients=42% females and 27% males),

possibly a reflection of breast cancer screening from the age of 47 years and over. Place of birth was largely incomplete, with outpatients having marginally poorer completion than inpatients (68% vs 70%); where completed most patients were born in the UK.

Table 8.3 : Dataset A and B-Inpatient and all outpatient, age and place of birth by gender

			Males		Females		Total	
Dataset	Factor	**Grouping**	**N**	**%**	**N**	**%**	**N**	**%**
A=Inpatients	Age	<=60	2315	24	3252	35	5567	30
		>60	7148	76	5948	65	13096	70
	Place of	UK	3114	33	2721	30	5835	31
	Birth	Non-UK	109	1	102	1	211	1
		Blank	6240	66	6377	69	12617	68
B=Outpatients	Age	<=60	2167	27	4361	42	6528	36
		>60	5740	73	6054	58	11794	64
	Place of	UK	2034	26	2310	22	4344	24
	Birth	Non-UK	70	1	88	1	158	1
		Blank	5803	73	8017	77	13820	75

A small proportion of in- and outpatients declined to provide data on religious beliefs (0.54% vs 0.36% respectively, see Table 8.4). Furthermore, large proportions in both in- and outpatients were coded as 'not specified' (37.16% and 41.23% respectively). Where reported, data on religion showed the majority of the population identified themselves as of Christian denomination, these were

combined into a new 'Christian all denominations' category (58.58% inpatients and 55.28% of outpatients).

Table 8.4 : Dataset A and B-Religious affiliation Inpatient and all outpatients

Factor	Grouping	A=Inpatients		B=Outpatients	
		N	%	N	%
Religious affiliation	Agnosticism	59	0.32	38	0.21
	Atheist	48	0.26	38	0.21
	Christian (all denominations)	10932	58.58	10129	55.28
	Declined to give	100	0.54	66	0.36
	Hinduism	154	0.83	144	0.79
	Islam	152	0.81	115	0.63
	Judaism	6	0.03	6	0.03
	Not Specified	6936	37.16	7554	41.23
	Other	18	0.10	13	0.07
	Satanism	-	-	1	0.01
	Sikhism	250	1.34	211	1.15
	Buddhism	8	0.04	7	0.04
	Total	18663	100	18322	100

Language was completed well with all patients having entries in both datasets (Table 8.5). The overwhelmingly majority (99%) spoke English. The remaining 1% comprised of Panjabi, Polish, Urdu and Guajarati amongst others. This revealed that the vast majority of Coventry's ethnic minority population had some level of English, which may be directly associated with country of birth and time in the UK. However, birthplace is poorly completed so I was not able to verify this (Table 8.3).

Table 8.5 : Dataset A and B-Inpatient and outpatient reported language

Factor	Grouping	A=Inpatients		B=Outpatients	
		N	%	N	%
Language	Albanian	1	0.01	1	0.01
	Amharic	1	0.01	1	0.01
	Arabic	1	0.01	1	0.01
	Bengali	5	0.03	2	0.01
	Bosnian	1	0.01	1	0.01
	British Sign Language	2	0.01	1	0.01
	Chinese	2	0.01	1	0.01
	Croatian	1	0.01	-	0.01
	Czech	3	0.02	2	0.01
	Dutch, Flemish	2	0.01	1	0.01
	English	18527	99.27	18228	99.49
	French	3	0.02	2	0.01
	German	2	0.01	1	0.01
	Greek	1	0.01	-	0.01
	Gujarati	10	0.05	9	0.05
	Hindi	5	0.03	4	0.02
	Italian	4	0.02	1	0.01
	Kurdish	1	0.01	3	0.02
	Latvian	1	0.01	1	0.01
	Lingala	1	0.01	1	0.01
	Malayalam	1	0.01	-	0.01
	Nepali	1	0.01	1	0.01
	Other	1	0.01	2	0.01
	Panjabi	39	0.21	26	0.14
	Persian (farsi)	-	-	1	0.01
	Polish	17	0.09	14	0.08
	Portuguese	3	0.02	3	0.02
	Romanian	1	0.01	1	0.01
	Russian	2	0.01	1	0.01
	Shona	1	0.01	1	0.01
	Tamil	2	0.01	1	0.01
	Thai	1	0.01	-	-
	Tigrinya	1	0.01	-	-
	Turkish	1	0.01	-	-
	Urdu	17	0.09	10	0.05
	Vietnamese	1	0.01	-	-
	Total	18663	100	18322	100

By categorising ICD 10 codes, I was able to code cancer sites and present them split by gender, see Table 8.6. The most common primary malignancies in the female subset were breast (24.9%), ill-defined malignancies (11.7%) and cancers of the female genital organs (8.7%). Prostate, digestive and colorectal and lung were the most common in males (20.3%, 11.6%, 9.9% and 9.9% respectively).

Table 8.6 : Dataset A-Inpatients by primary diagnosis and gender

Disease site	Females		Males		Total	
	N	**%**	**N**	**%**	**N**	**%**
Head and Neck	239	2.6	547	5.8	786	4.2
Bladder	261	2.8	800	8.5	1061	5.7
Brain/Eye	215	2.3	375	4	590	3.2
Breast	2294	24.9	18	0.2	2312	12.4
Colorectal	776	8.4	935	9.9	1711	9.2
Digestive	720	7.8	1093	11.6	1813	9.7
Female genital organs	803	8.7	-	-	803	4.3
Ill defined	1073	11.7	904	9.6	1977	10.6
Leukaemia	268	2.9	372	3.9	640	3.4
Lung	764	8.3	939	9.9	1703	9.1
Lymphoma.	420	4.6	485	5.1	905	4.9
Male genital organs	-	-	191	2	191	1
Melanoma	126	1.4	147	1.6	273	1.5
Mesothelioma	80	0.9	120	1.3	200	1
Myeloma	196	2.1	217	2.3	413	2.2
Ovary	550	6	-	-	550	3
Prostate	-	-	1917	20.3	1917	10.3
Thyroid	201	2.2	79	0.8	280	1.5
Urinary tract	185	2	279	2.6	464	2
Other	29	0.3	45	0.5	74	0.4
Total	9200		9463		18663	100

When examining all diagnoses using the transposed copy of dataset A, cancers of ill-defined origin were the most common diagnosis over the six year period (30.2%), followed by breast cancer (19.9%), with lung cancer overtaking malignancies of the females genital organs (6.9% vs 6.7%), see Table 8.7. Similarly in males, ill-defined malignancies were the most common cancer over the six year period, followed by prostate and the addition of malignancies of the digestive organs (25.3%, 17.2% and 9.3% respectively).

Table 8.7 Dataset A-Inpatients, all diagnoses by gender (multiple patients from transposed dataset)

Disease site	Females		Males		Total	
	N	**%**	**N**	**%**	**N**	**%**
Head and Neck	245	2	563	4.7	808	3.3
Bladder	267	2.2	828	6.9	1095	4.5
Brain/Eye	222	1.8	382	3.2	604	2.5
Breast	2427	19.9	21	0.2	2448	10.1
Colorectal	816	6.7	980	8.1	1796	7.4
Digestive	734	6	1123	9.3	1857	7.7
Female genital organs	817	6.7	-	-	817	3.4
Ill defined	3677	30.2	3042	25.3	6719	27.7
Leukaemia	269	2.2	384	3.2	653	2.7
Lung	842	6.9	1032	8.6	1874	7.7
Lymphoma.	432	3.6	497	4.1	929	3.8
Male genital organs	-	-	193	1.6	193	0.8
Melanoma	132	1.1	156	1.3	288	1.2
Mesothelioma	88	0.7	121	1	209	0.9
Myeloma	200	1.6	221	1.8	421	1.7
Ovary	584	4.8	-	-	584	2.4
Prostate	-	-	2067	17.2	2067	8.5
Thyroid	204	1.7	83	0.7	287	1.2
Urinary tract	198	1.6	304	2.5	502	2.1
Other	32	0.3	46	0.4	78	0.3
Total	12186	50.3	12043	49.7	24229	100

8.3.4 C=Overlapped dataset (in- and outpatients)

10,251 patients were treated as both in- and outpatients. Ethnicity data were compared across the sets to check for agreement. Ethnicity was the same in 10,215 of 10,251 patients (99.6%). In the 36 patients where there was disagreement, many were due to unknown codes recorded in one dataset, e.g. 12 patients were coded as not known and 12 as not stated in outpatients but were coded as White British in the inpatient set (Table 8.8). A direct conflict of codes was detected in only three patients (in bold in Table 8.8). One patient was coded as an Irish inpatient but White British as an outpatient. Two other patients were coded as Indian as inpatients but Mixed in outpatients (possible error as Asian not selected from the Mixed categories). As inpatient data were generally better completed, I was more confident in the quality so took this to be the more accurate of the two datasets.

Table 8.8 Dataset C-Inpatients and outpatient ethnicity comparison results for 36 patients with disagreement

A=Inpatients	B=Outpatients	N
White British	Not known	12
	Not stated	12
Irish	**White British**	**1**
	Not known	1
	Not stated	1
Other White	Not known	1
	Not stated	1
Indian	**Mix-White & Caribbean**	**1**
	Mix Other	**1**
	Not stated	2
Not known	White British	1
Not stated	White British	2
Total		**36**

8.3.5 D=Final dataset (all inpatients plus any additional outpatients)

Ethnic groups were combined to eliminate groups with small numbers for the final analyses. The smallest groups with recorded ethnicity were those who chose not to specify, all mixed categories, Chinese, Bangladeshi, Pakistani, Other Asian, African and Other Black. For this reason, I combined Pakistani, Bangladeshi and Other Asian background together to form a new 'Other Asian' category and African and Other Black background into a 'Other Black' category, as well combining Chinese into the already existing 'Other Ethnic' (see Table 8.9). I also created a new 'All Mixed' category for all mixed groups.

Table 8.9 Dataset D-Final ethnic group

Factor	Grouping	n	%
Ethnic group	A=White British	21570	81
	B=Irish	580	2
	C=Other White	390	1
	D-G=All Mixed ethnicities	78	0.3
	H=Indian	659	2
	J-L=Pakistani, Bangladeshi and Other Asian	219	0.8
	M=Caribbean	163	0.6
	N and P=African and Other Black	131	0.5
	S=Other ethnic	213	0.8
	Unknown ethnicity	2731	10
	Total	26734	100

Due to small numbers I also amalgamated cancer codes as shown below:

1. Digestive organs: small intestine, oesophagus, stomach
2. Female genital organs: Uterus, placenta, Vulva
3. Male genital organs: Testis, penis
4. Lymphoma: Hodgkins, Non-Hodgkins, Other Lymphoma

Age and gender information was available for all 26,734 patients in the final set (Table 8.10). There were more females (53%) than males (47%) overall. The proportion of females aged 65 years and under was higher in comparison to males (38% vs 24% respectively). Place of birth was non-informative given the large proportion of missing data.

Table 8.10 Dataset D-Final age group and place of birth by gender

		Females (n=14103)		Males (n=12631)		Total (n=26734)	
Factor	**Grouping**	**N**	**%**	**N**	**%**	**N**	**%**
Age group	<=65 years	5308	38	3081	24	8389	31
	>65 years	8795	62	9550	76	18345	69
Place of birth	UK	3322	24	3514	28	6836	26
	Non-UK	127	0.9	125	1	252	0.9
	Missing	5	0.04	5	0.04	10	0.04
	Not known/stated	10649	76	8987	71	19636	73

Forty per cent of patients did not specify any religious beliefs with a further 0.4% declining to provide this information (Table 8.11). 56% of patients were Christian (56%) with other religions accounting for 3.7%. Sikhism, Hinduism and Islam (1.2%, 0.8% and 0.7% respectively) were the largest other groups in this population.

Table 8.11 Dataset D-Final religion

Religion	N	%
Christian (all denominations)	14920	56
Not Specified	10825	40
Sikhism	319	1.2
Hinduism	206	0.8
Islam	186	0.7
Declined to give	113	0.4
Agnosticism	64	0.2
Atheist	58	0.2
Other	23	0.1
Buddhism	10	0.04
Judaism	9	0.03
Satanism	1	0.004
Total	26734	100

An overwhelming majority of patients spoke English (99.4%) as shown in Table 8.12. The three most common languages other than English were Panjabi (0.18%), Polish (22%) and Urdu (0.07%). Information regarding proficiency was not available.

Table 8.12 Dataset D-Final language

Factor	Grouping	N	%
Language	Albanian	1	0.004
	Amharic	1	0.004
	Arabic	1	0.004
	Bengali	5	0.02
	Bosnian	1	0.004
	British Sign Language	2	0.01
	Chinese	2	0.01
	Croatian	1	0.004
	Czech	3	0.01
	Dutch, Flemish	2	0.01
	English	26574	99.4
	French	3	0.01
	German	2	0.01
	Greek	1	0.004
	Gujarati	13	0.05
	Hindi	6	0.02
	Italian	5	0.02
	Kurdish	3	0.01
	Latvian	1	0.004
	Lingala	1	0.004
	Malayalam	1	0.004
	Nepali	1	0.004
	Other	2	0.01
	Panjabi	47	0.18
	Persian (Farsi)	1	0.004
	Polish	22	0.08
	Portuguese	3	0.01
	Romanian	1	0.004
	Russian	3	0.01
	Shona	1	0.004
	Tamil	2	0.01
	Thai	1	0.004
	Tigrinya	1	0.004
	Turkish	1	0.004
	Urdu	18	0.07
	Vietnamese	1	0.004
	Total	26734	100

8.3.6 Primary cancer diagnosis by ethnic group (dataset D)

ICD codes were only recorded for inpatient; disease site was unknown for the 8,071 patients treated as outpatients only. Table 8.13 below shows the rates of cancer for the final dataset (daatset D). Data were not also presented split by gender due to the small number of ethnic minorities. Outpatients were included as a separate row.

Tables 8.13 and 8.14 show figures for primary cancer site by ethnic group over the six year period (2007-2012). Figures for 2012 have also been presented in order to allow comparability between the HES and CRN recruitment data. Table 8.13 presents column per cents which highlights the distribution of cancers within each ethnic group, whereas Table 8.14 presents row per cents to show the distribution of ethnic minorities across disease site.

Overall, the numbers of ethnic minority patients was small. The proportion of patients with not known/not specified ethnicity data accounted for 10% of the six year data. However, looking at this category over the six year period it is possible to see a statistically significant trend showing an increase in unknown ethnicity codes from 5% in 2007 to 13% three years later in 2010 and 15% in 2012 (Mantel-Haenszel Chi-Square $p<0.0001$). This may be a reflection of the increased strain on NHS resources.

Based upon the available data it was possible to see that breast cancer was the most prevalent malignancy in six ethnic groups, White British (9%), Other White

(9.5%), Indian (13.5%), Pakistani/Bangladeshi/other Asians (11.9%), other ethnic group (11.3%) and unknown ethnicity (3.7%, see Table 8.13). Prostate cancer was the most common in three groups including Irish (12.6%), Caribbean (24.5%) and the new All mixed category (15.4%). The high proportion of prostate cancers in the Caribbean group confirms existing evidence of higher incidence in this population (National Cancer Intelligence Network, 2009). Interestingly, the high proportion of prostate cancers in the Mixed group appeared to be due to the mixing of the African or Caribbean population with other ethnic groups ('White and Caribbean', 'White and African' and 'White and Other' patients in this dataset). Lymphoma was the most common in the African and Other Black group (10.7%).

Table 8.14 shows White British patients unsurprisingly accounting for the bulk of patients in every disease site ranging from the lower 74.4% in cancers of the male genitals to a high of 90.5% in melanoma. Looking at ethnic minority disease aside from the White British population, the Irish group were shown to be the most afflicted by bladder (3%), lung (3.3%) and prostate (3.8%) cancers. The highest ethnic minority rates of eye/brain cancers (2.4%), leukaemia (3.6%), mesothelium (4.5%), myeloma (4.8%), thyroid (4.3%) and other malignancies (5.4%) were in the Indian group. However, sites such as head and neck (3.2%), breast (4.4%), colorectal (3.8%), digestive (3.6%), malignancies of the female genitals (9.3%), ill-defined cancers (3.1%), lymphoma (3.4%), male genitals (17.3%), melanoma (5.1%) and prostate (10%) were highest in the unknown ethnicity group.

As the inpatient dataset contained multiple diagnoses it was important to explore all malignancies. Tables 8.15 and 8.16 are in the same format as above but report all cancers suffered. Ill-defined malignancies were the most common across all ethnic groups with the exception of Caribbean patients where prostate cancer was slightly higher accounting for 25.2% (Table 8.15). The distribution of all malignancies across ethnic minorities group (i.e. excluding White British) was similar to that of the primary diagnosis. The 'unknown ethnicity' group suffered the largest proportion of disease across the majority of disease sites with the exception of eye/brain, leukaemia, lymphoma, mesothelium, myeloma, and other cancers where the Indian group were shown to carry greater burden (2.4, 3.6, 4, 4.5, 4.8 and 5.4% respectively. The Irish population were seen to suffer the most bladder cancers (3%), lung cancers (3.3%) and prostate cancers (3.8%, see Table 8.16). Cancer rates in ethnic minority populations remained small even when considering all malignancies.

Table 8.13 Dataset D-Primary cancer diagnosis by ethnic group (column %)

	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnicity	Unknown ethnicity	Total
Head & Neck	692 (3)	12 (2)	14 (3.6)	3 (3.9)	22 (3.3)	9 (4.1)	3 (1.8)	1 (0.8)	5 (2.4)	25 (0.9)	786 (2.9)
Bladder	952 (4)	32 (5.5)	18 (4.7)	1 (1.3)	24 (3.6)	12 (5.5)	5 (3)	1 (0.8)	6 (2.8)	10 (0.4)	1061 (4.0)
Breast	1959 (9)	44 (7.6)	37 (9.5)	9 (11.5)	89 (13.5)	26 (11.9)	9 (5.5)	13 (9.9)	24 (11.3)	102 (3.7)	2312 (8.6)
Colorectal	1500 (7)	44 (7.6)	29 (7.4)	2 (2.6)	30 (4.6)	11 (5)	9 (5.5)	6 (4.6)	15 (7)	65 (2.4)	1711 (6.4)
Digestive	1578 (7)	38 (6.6)	21 (5.4)	3 (3.9)	46 (7)	22 (10)	16 (9.8)	10 (7.6)	13 (6.1)	66 (2.4)	1813 (6.8)
Eye/Brain	530 (2)	6 (1)	8 (2)	1 (1.3)	14 (2.1)	8 (3.7)	2 (1.2)	2 (1.5)	6 (2.3)	13 (0.5)	590 (2.2)
Female genitals	647 (3)	10 (1.7)	18 (4.6)	3 (3.9)	24 (3.6)	4 (1.8)	6 (3.7)	5 (3.8)	11 (5.2)	75 (2.6)	803 (3.0)
Ill defined	1751 (8)	36 (6.2)	32 (8.2)	4 (5.1)	39 (5.9)	12 (5.5)	12(7.4)	11 (8.4)	18 (8.5)	62 (2.3)	1977 (7.4)
Leukaemia	543 (3)	19 (3.3)	6 (1.5)	1 (1.3)	23 (3.5)	13 (5.9)	2 (1.2)	9 (6.9)	9 (4.2)	15 (0.6)	640 (2.4)
Lung	1503 (7)	56 (9.7)	28 (7.2)	4 (5.1)	21 (3.2)	13 (5.9)	5 (3.1)	4 (3)	16 (7.5)	53 (1.9)	1703 (6.4)
Lymphoma	751 (3)	23 (4)	17 (4.4)	1 (1.3)	37 (5.6)	16 (7.3)	6 (3.7)	14 (10.7)	9 (4.2)	31 (1.1)	905 (3.4)
Male genitals	142 (0.7)	1 (0.2)	5 (1.3)	0	6 (0.9)	3 (1.4)	0	0	1 (0.5)	33 (1.2)	191 (0.7)
Melanoma	247 (1.2)	5 (0.9)	4 (1)	0	2 (0.3)	0	0	0	1 (0.5)	14 (0.5)	273 (1.0)
Mesothelium	171 (0.8)	3 (0.5)	3 (0.8)	1 (1.3)	9 (1.4)	0	2 (1.2)	3 (2.3)	4 (1.9)	4 (0.2)	200 (0.7)
Myeloma	343 (1.6)	8 (1.4)	10 (2.6)	2 (2.6)	20 (3)	2 (0.9)	8 (4.9)	5 (3.8)	3 (1.4)	12 (0.4)	413 (1.5)
Ovary	447 (2)	9 (1.6)	8 (2.1)	4 (5.1)	16 (2.4)	4 (1.8)	2 (1.2)	0	5 (2.4)	55 (2)	550 (2.1)
Prostate	1631 (7.6)	73 (12.6)	28 (7.2)	12 (15.4)	62 (9.4)	16 (7.3)	40 (24.5)	13 (9.9)	10 (4.7)	32 (1.2)	1917 (7.2)
Thyroid	219 (1)	3 (0.5)	10 (2.6)	4 (5.1)	12 (1.8)	5 (2.3)	0	5 (3.8)	6 (2.8)	16 (0.6)	280 (1.0)
Urinary tract	396 (1.8)	14 (2.4)	9 (2.3)	1 (1.3)	14 (2.1)	8 (3.7)	3 (1.8)	3 (2.3)	3 (1.4)	13 (0.5)	464 (1.7)
Other	63 (0.3)	0	3 (0.8)	0	4 (0.6)	0	1 (0.6)	1 (0.8)	1 (0.5)	1 (0.04)	74 (0.3)
Outpatients	5505 (25.5)	144 (24.8)	82 (21)	22 (28.2)	145 (22)	35 (16)	32 (19.6)	25 (19.1)	47 (22.1)	2034 (74.5)	8071 (30.2)
2007-12	21570	580	390	78	659	219	163	131	213	2731	26734 (100)
2012 only	3205 (76)	73 (2)	81 (2)	10 (0.2)	105 (3)	35 (0.8)	26 (0.6)	20 (0.5)	29 (0.7)	629 (15)	4213 (100)

Table 8.14 : Dataset D-Primary cancer diagnosis by ethnic group (row %)

	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnicity	Unknown ethnicity	Total
Head & Neck	692 (88)	12 (1.5)	14 (1.8)	3 (0.4)	22 (2.8)	9 (1.2)	3 (0.4)	1 (0.1)	5 (0.6)	25 (3.2)	786
Bladder	952 (89.7)	32 (3)	18 (1.7)	1 (0.1)	24 (2.3)	12 (1.1)	5 (0.5)	1 (0.1)	6 (0.6)	10 (0.9)	1061
Breast	1959 (84.7)	44 (1.9)	37 (1.6)	9 (0.4)	89 (3.9)	26 (1.1)	9 (0.4)	13 (0.6)	24 (1)	102 (4.4)	2312
Colorectal	1500 (87.7)	44 (2.6)	29 (1.7)	2 (0.1)	30 (1.8)	11 (0.6)	9 (0.5)	6 (0.4)	15 (0.9)	65 (3.8)	1711
Digestive	1578 (87)	38 (2.1)	21 (1.2)	3 (0.2)	46 (2.5)	22 (1.2)	16 (0.9)	10 (0.6)	13 (0.7)	66 (3.6)	1813
Eye/Brain	530 (89.8)	6 (1)	8 (1.4)	1 (0.2)	14 (2.4)	8 (1.4)	2 (0.3)	2 (0.3)	6 (1)	13 (2.2)	590
Female genitals	647 (80.1)	10 (1.3)	18 (2.2)	3 (0.4)	24 (3)	4 (0.5)	6 (0.8)	5 (0.6)	11 (1.4)	75 (9.3)	803
Ill defined	1751 (88.6)	36 (1.8)	32 (1.6)	4 (0.2)	39 (2)	12 (0.6)	12(0.6)	11 (0.6)	18 (0.9)	62 (3.1)	1977
Leukaemia	543 (84.8)	19 (3)	6 (0.9)	1 (0.2)	23 (3.6)	13 (2)	2 (0.3)	9 (1.4)	9 (1.4)	15 (2.3)	640
Lung	1503 (88.3)	56 (3.3)	28 (1.6)	4 (0.2)	21 (1.2)	13 (0.8)	5 (0.3)	4 (0.2)	16 (0.9)	53 (3.1)	1703
Lymphoma	751 (83)	23 (2.5)	17 (1.9)	1 (0.1)	37 (4)	16 (1.8)	6 (0.7)	14 (1.6)	9 (1)	31 (3.4)	905
Male genitals	142 (74.4)	1 (0.52)	5 (2.6)	0	6 (3)	3 (1.6)	0	0	1 (0.5)	33 (17.3)	191
Melanoma	247 (90.5)	5 (1.8)	4 (1.5)	0	2 (0.7)	0	0	0	1 (0.4)	14 (5.1)	273
Mesothelium	171 (85.5)	3 (1.5)	3 (1.5)	1 (0.5)	9 (4.5)	0	2 (1)	3 (1.5)	4 (2)	4 (2)	200
Myeloma	343 (83.1)	8 (1.9)	10 (2.4)	2 (0.5)	20 (4.8)	2 (0.5)	8 (1.9)	5 (1.2)	3 (0.7)	12 (2.9)	413
Ovary	447 (81.3)	9 (1.6)	8 (1.5)	4 (0.7)	16 (2.9)	4 (0.7)	2 (0.4)	0	5 (0.9)	55 (10)	550
Prostate	1631 (85.1)	73 (3.8)	28 (1.5)	12 (0.6)	62 (3.2)	16 (0.8)	40 (2.1)	13 (0.7)	10 (0.5)	32 (1.7)	1917
Thyroid	219 (78.2)	3 (1.1)	10 (3.6)	4 (1.4)	12 (4.3)	5 (1.8)	0	5 (1.8)	6 (2)	16 (5.7)	280
Urinary tract	396 (85.3)	14 (3)	9 (1.9)	1 (0.2)	14 (3)	8 (1.7)	3 (0.7)	3 (0.7)	3 (0.7)	13 (2.8)	464
Other	63 (85.1)	0	3 (4)	0	4 (5.4)	0	1 (1.4)	1 (1.4)	1 (1.4)	1 (1.4)	74
Outpatients	5505 (68.2)	144 (1.8)	82 (1)	22 (0.3)	145 (1.8)	35 (0.4)	32 (0.4)	25 (0.3)	47 (0.6)	2034 (25.2)	8071
2007-12	21570 (81)	580 (2.2)	390 (1.5)	78 (0.3)	659 (2.5)	219 (0.8)	163 (0.6)	131 (0.5)	213 (0.8)	2731 (10)	26734
2012 only	3205 (76)	73 (2)	81 (2)	10 (0.2)	105 (3)	35 (0.8)	26 (0.6)	20 (0.5)	29 (0.7)	629 (15)	4213 (100)

Table 8.15 Dataset D-All cancer diagnoses by ethnic group (multiple rows per patient, column %)

	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnicity	Unknown ethnicity	Total
Head & Neck	712 (3.4)	12 (2,2)	15 (3.8)	3 (4.2)	23 (3.6)	9 (3.8)	3 (1.9)	1 (0.8)	5 (2.5)	25 (2.7)	808 (3.3)
Bladder	982 (4.7)	33 (6)	18 (4.6)	1 (1.4)	24 (3.7)	12 (5)	5 (3.1)	1 (0.8)	6 (2.7)	13 (1.4)	1095 (4.5)
Breast	2081 (10)	45 (8.2)	38 (9.6)	10 (14.1)	94 (14.6)	26 (10.8)	9 (5.7)	13 (9.9)	25 (12.3)	107 (11.6)	2448 (10.1)
Colorectal	1572 (7,5)	47 (8.6)	33 (8.4)	2 (2.8)	32 (5)	12 (5)	9 (5.7)	6 (4.6)	15 (7.4)	68 (7.4)	1796 (7.4)
Digestive	1619 (7.7)	38 (7)	21(5.3)	4 (5.6)	46 (7)	23 (9.6)	17 (10.7)	10 (7.6)	13 (6.4)	66 (7.2)	1857 (7.7)
Eye/Brain	541 (2.6)	6 (1.1)	9 (2.3)	1 (1.4)	14 (2.2)	8 (3.3)	2 (1.3)	3 (2.3)	6 (3)	14 (1.5)	604 (2.5)
Female genitals	658 (3.2)	10 (1.8)	19 (4.8)	3 (4.2)	24 (3.4)	4 (1.7)	6 (3.8)	5 (3.8)	11 (5.4)	77 (8.4)	817 (3.4)
Ill defined	5870 (28)	135 (24.6)	103 (26)	17 (23.9)	153 (24)	63 (26.3)	38 (24)	32 (24.4)	51 (25)	257 (28)	6719 (27.7)
Leukaemia	554 (2.7)	19 (3.5)	7 ()1.8	1 (1.4)	24 (3.7)	13 (5.4)	2 (1.3)	9 (6.9)	9 (4.4)	15 (1.6)	653 (2.7)
Lung	1654 (8)	59 (10.7)	30 (7.6)	4 (5.6)	23 (3.6)	14 (5.8)	6 (3.8)	5 (3.8)	17 (8.4)	62 (6.4)	1874 (7.7)
Lymphoma	773 (3.7)	23 (4.2)	18 (4.6)	1 (1.4)	37 (5.8)	16 (6.7)	6 (3.8)	14 (10.7)	9 (4.4)	32 (3.5)	929 (3.8)
Male genitals	144 (0.7)	1 (0.2)	5 (1.3)	0	6 (0.9)	3 (1.3)	0	0	1 (0.5)	33 (3.6)	193 (0.8)
Melanoma	262 (1.3)	5 (0.9)	4 (1)	0	2 (0.3)	0	0	0	1 (0.5)	14 (1.5)	288 (1.2)
Mesothelium	180 (0.9)	3 (0.6)	3 (0.8)	1 (1.4)	9 (1.4)	0	2 (1.3)	3 (2.3)	4 (2)	4 (0.4)	209 (0.9)
Myeloma	351 (1.7)	8 (1.5)	10 (2.5)	2 (2.8)	20 (3.1)	2 (0.8)	8 (5.0)	5 (3.8)	3 (1.5)	12 (1.3)	421 (1.7)
Ovary	478 (2.3)	10 (1.8)	8 (2)	4 (5.6)	16 (2.5)	4 (1.7)	2 (1.3)	0	5 (2.5)	57 (6.2)	584 (2.4)
Prostate	1765 (8.4)	79 (14.4)	30 (7.6)	12 (16.9)	65 (10.1)	16 (6.7)	40 (25.2)	14 (10.7)	11 (5.4)	35 (3.8)	2067 (8.5)
Thyroid	224 (1.1)	3 (0.6)	11 (2.8)	4 (5.6)	12 (1.9)	6 (2.5)	0	5 (3.8)	6 (3)	16 (1.7)	287 (1.2)
Urinary tract	431 (2.1)	14 (2.6)	9 (2.3)	1 (1.4)	15 (2.3)	8 (3.3)	3 (1.9)	4 (3)	4 (2)	13 (1.4)	502 (2.1)
Other	66 (0.3)	0	3 (0.8)	0	4 (0.6)	1 (0.4)	1 (0.6)	1 (0.8)	1 (0.5)	1 (0.1)	78 (0.3)
Total 2007-12	20917 (86.3)	550 (2.3)	394 (1.6)	71 (0.3)	643 (2.7)	240 (1.0)	159 (0.7)	131 (0.5)	203 (0.8)	921 (3.8)	24229 (100)

Table 8.16 Dataset D-All cancer diagnoses by ethnic group (multiple rows per patient, row%)

	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnicity	Unknown ethnicity	Total
Head & Neck	712 (88.1)	12 (1.5)	15 (1.9)	3 (0.4)	23 (2.9)	9 (1.1)	3 (0.4)	1 (0.1)	5 (0.6)	25 (3.1)	808 (3.3)
Bladder	982 (89.7)	33 (3)	18 (1.6)	1 (0.1)	24 (2.2)	12 (1.1)	5 (0.5)	1 (0.1)	6 (0.6)	13 (1.2)	1095 (4.5)
Breast	2081 (85)	45 (1.8)	38 (1.6)	10 (0.4)	94 (3.8)	26 (1.1)	9 (0.4)	13 (0.6)	25(1)	107 (4.4)	2448 (10.1)
Colorectal	1572 (87.5)	47 (2.6)	33 (1.8)	2 (0.1)	32 (1.8)	12 (0.7)	9 (0.5)	6 (0.3)	15 (0.8)	68 (3.8)	1796 (7.4)
Digestive	1619 (87.2)	38 (2.1)	21(1.1)	4 (0.2)	46 (2.5)	23 (1.2)	17 (0.9)	10 (0.5)	13 (0.7)	66 (3.6)	1857 (7.7)
Eye/Brain	541 (89.6)	6 (1)	9 (1.5)	1 (0.2)	14 (2.3)	8 (1.3)	2 (0.3)	3 (0.5)	6 (1)	14 (2.3)	604 (2.5)
Female genitals	658 (80.5)	10 (1.2)	19 (2.3)	3 (0.4)	24 (2.9)	4 (0.5)	6 (0.7)	5 (0.6)	11 (1.4)	77 (9.4)	817 (3.4)
Ill defined	5870 (87.4)	135 (2)	103 (1.5)	17 (0.3)	153 (2.3)	63 (0.9)	38 (0.6)	32 (0.5)	51 (0.8)	257 (3.8)	6719 (27.7)
Leukaemia	554 (84.8)	19 (2.9)	7 (1.1)	1 (0.2)	24 (3.7)	13 (2.0)	2 (0.3)	9 (1.4)	9 (1.4)	15 (2.3)	653 (2.7)
Lung	1654 (88.3)	59 (3.2)	30 (1.6)	4 (0.2)	23 (1.2)	14 (0.8)	6 (0.3)	5 (0.3)	17 (0.9)	62 (3.3)	1874 (7.7)
Lymphoma	773 (83.2)	23 (2.5)	18 (1.9)	1 (0.1)	37 (4.0)	16 (1.7)	6 (0.7)	14 (1.5)	9 (1)	32 (3.4)	929 (3.8)
Male genitals	144 (75)	1 (0.5)	5 (2.6)	0	6 (3.1)	3 (1.6)	0	0	1 (0.5)	33 (17.1)	193 (0.8)
Melanoma	262 (91)	5 (1.7)	4 (1.4)	0	2 (0.7)	0	0	0	1 (0.4)	14 (4.9)	288 (1.2)
Mesothelium	180 (86)	3 (1.4)	3 (1.4)	1 (0.5)	9 (4.3)	0	2 (1.0)	3 (1.4)	4 (1.9)	4 (1.9)	209 (0.9)
Myeloma	351 (83.4)	8 (1.9)	10 (2.4)	2 (0.5)	20 (4.8)	2 (0.5)	8 (1.9)	5 (1.2)	3 (0.7)	12 (2.6)	421 (1.7)
Ovary	478 (82)	10 (1.7)	8 (1.4)	4 (0.7)	16 (2.7)	4 (0.7)	2 (0.3)	0	5 (0.9)	57 (9.8)	584 (2.4)
Prostate	1765 (85.4)	79 (3.8)	30 (1.4)	12 (0.6)	65 (3.1)	16 (0.8)	40 (1.9)	14 (0.7)	11 (0.5)	35 (1.7)	2067 (8.5)
Thyroid	224 (78.1)	3 (1)	11 (0.2)	4 (1.4)	12 (4.2)	6 (2.1)	0	5 (1.7)	6 (2.1)	16 (5.6)	287 (1.2)
Urinary tract	431 (86)	14 (2.8)	9 (1.8)	1 (0.2)	15 (3.0)	8 (1.6)	3 (0.6)	4 (0.8)	4 (0.8)	13 (2.6)	502 (2.1)
Other	66 (85)	0	3 (0.1)	0	4 (5.1)	1 (1.3)	1 (1.3)	1 (1.3)	1 (1.3)	1 (1.3)	78 (0.3)
2007-12	20917 (86)	550 (2.3)	394 (1.6)	71 (0.3)	643 (2.6)	240 (1.0)	159 (0.7)	131 (0.5)	203 (0.8)	921 (3.8)	24229

8.3.7 Activity type

Inpatients were admitted for either elective or non-elective treatment. Elective included patients being treated as day cases as well as those staying overnight. Overall, 66% of patients were admitted to hospital for elective treatment. Table 8.17 below shows no differences in the type of activity by ethnic group over the six year period. The proportion of patients having elective surgery was highest in the unknown ethnicity category (81%). For the remaining ethnic groups, the numbers having elective treatment were similar. Table 8.18 shows the same to be true when examining data for 2012 independently, where 68% were admitted for elective treatment. Any variations can be explained by the small sample.

Table 8.17 Activity type by ethnic group 2007-2012

Activity type	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnic	Unknown ethnicity	Total
Elective	10487 (65)	281 (64)	197 (64)	39 (70)	331 (64)	101 (55)	77 (59)	66 (62)	100 (60)	563 (81)	12242 (66)
Non-Elective	5578 (35)	155 (36)	111 (36)	17 (30)	183 (36)	83 (45)	54 (41)	40 (38)	66 (40)	134 (19)	6421 (44)
2007-12	16065	436	308	56	514	184	131	106	166	697	18663

Table 8.18 Activity type by ethnic group for 2012

Activity type	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnic	Unknown Ethnicity	Total
Elective	1785 (68)	37 (67)	52 (74)	6 (67)	61 (67)	13 (43)	18(72)	13 (87)	18 (75)	118(81)	2121 (68)
Non-Elective	852 (32)	18 (33)	18 (26)	3(33)	30(33)	17 (57)	7(28)	2 (13)	6 (25)	27(19)	980 (32)
2012 only	2637	55	70	9	91	30	25	15	24	145	3101

8.3.8 Surgery

Of the 18,663 patients admitted to hospital 13,713 (73%) did not have surgery (Table 8.19). The most common treatments for these non-surgery patients were general medicine (28%), clinical oncology (18%) and urology (16%). In the 4,950 (27%) that had surgery, 31% were coded as general surgery and 28% as having breast surgery. No differences in surgery were detected by ethnic group when examining the six year data or 2012 data (Table 8.20). Again any fluctuations can be explained by the small samples.

Table 8.19 Surgery by ethnic group 2007-2012

Surgery	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnic	Unknown ethnicity	Total
Yes	4335 (27)	110 (25)	80 (26)	13 (23)	146 (28)	45 (24)	26 (20)	22 (21)	48 (29)	125 (18)	4950 (27)
No	11730 (73)	326 (75)	228 (74)	43 (77)	368(72)	139(76)	105 (80)	84 (79)	118 (71)	572(82)	13713 (73)
2007-12	16065	436	308	56	514	184	131	106	166	697	18663

Table 8.20 Surgery by ethnic group 2012

Surgery	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnic	Unknown ethnicity	Total
Yes	708 (27)	14 (25)	26 (37)	1(11)	20(22)	7 (23)	3 (12)	3 (20)	7 (29)	19 (13)	808 (26)
No	1929 (73)	41 (75)	44 (63)	8 (89)	71 (78)	23 (77)	22 (88)	12 (80)	17 (71)	126 (87)	2293(74)
2012	2637	55	70	9	91	30	25	15	24	145	3101

8.3.9 Episode outcome

Episode outcome data were missing for three patients (two being White British and one of Indian origin) bringing the total down to 18,660. The majority of inpatients were discharged on clinical advice (94%, Table 8.21). A small proportion self-discharged or were discharged by a relative (0.2%), the remaining 6% died whist in hospital. The data were very similar across the board with no differences detected by ethnicity. The picture was very similar when examining 2012 data independently (Table 8.22). Based upon these data, there was no evidence to show a difference in activity type or outcome by ethnic group. However, it is important to note the large proportion of missing ethnicity data, the unknown ethnicity group was the second largest preceded only by the White British group.

To summarise, the data shown above were too limited to permit any conclusions. The high proportion of missing data limits interpretation further as we do not know if the data were missing at random across all ethnic groups or for restricted to particular ethnic groups, e.g. non-English speaking.

Table 8.21 Episode outcome by ethnic group 2007-2012

Activity type	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnic	Unknown ethnicity	Total
Self/ relative	29 (0.2)	0	0	0	1 (0.2)	0	0	0	0	1 (0.1)	31 (0.2)
On clinical advice	15021 (94)	404 (93)	292 (95)	53 (95)	483 (94)	177 (96)	126 (96)	101 (95)	153 (92)	653 (94)	17463 (94)
Patient died	1013 (6)	32 (7)	16 (5)	3 (5)	29 (6)	7 (4)	5 (4)	5 (5)	13 (8)	43 (6)	1166 (6)
2007-12	16063	436	308	56	513	184	131	106	166	697	18660

Table 8.22 Episode outcome by ethnic group 2012

Activity type	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnic	Unknown ethnicity	Total
Self/ relative	6 (0.2)	0	0	0	0	0	0	0	0	1 (0.7)	7 (0.2)
On clinical advice	2514 (94)	53(96)	68(97)	8 (89)	89 (98)	29 (97)	24 (96)	15 (100)	23 (96)	141(97)	2964 (96)
Patient died	117 (5)	2 (4)	2 (3)	1 (11)	2 (2)	1 (3)	1 (4)	0	1 (4)	3 (2)	130 (4)
2012	2637	55	70	9	91	30	25	15	24	145	3101

8.3.10 Representation of ethnic minorities in clinical trials

In order to assess the representation of ethnic minorities in clinical trials on a local level I used data from HES as presented above and data from a cancer research network. The total number of patients identified by healthcare professionals as potential participants in clinical trials across the entire cancer network was 1271 for 2012/2013 spanning across five trusts. Of the 1271, 897 were treated at the large local teaching hospital, with 618 (69%) of these being successfully recruited to a trial (Table 8.23). On the whole, only a small proportion were not invited to participate due to clinician decision (4%). At the time of the analysis 7% of patients were still being screened and a further 20% had refused to participate. Ethnicity was not known in 15% of the 897 patients identified at the local hospital.

Only the ethnic groups of White British, Indian and unknown ethnicity had counts greater than 20. Similar proportions of patients from these three groups were recruited into clinical trials (69%, 71% and 63% respectively). Early indications suggest that ethnic minority patients may be more likely to decline participation (e.g. 33% Mixed, 29% Indian, 40% Other Asian, 33% Other Ethnicity group and 28% Not known) compared to white groups (18% White British, 15% Irish and 19% other White).

HES data from 2012 was used to form a crude denominator to estimate the representativeness of ethnic minority patients in trials. However, patients having

surgery, dying whilst in hospital and those self-discharging or being discharged by a relative were deducted from the denominator as shown in Table 8.24.

Table 8.23: Trial status for potential participants identified by the Cancer Research Network

	White British	Irish	Other White	All Mixed	Indian	Other Asian	All Black	Other Ethnic	Unknown ethnicity	Total
Clinician decision	47 (7)	-	3 (19)	1 (33)	2 (10)	1 (10)	1 (12)	-	8 (6)	63 (4)
Patient declined	105 (15)	2 (15)		1 (33)	4 (19)	3 (30)	-	2 (33)	32 (24)	149 (20)
Screened	59 (9)	-	-	-	-	-	-	-	8 (6)	67 (7)
Recruited	477 (69)	11 (85)	13 (81)	1 (33)	15 (71)	6 (60)	7 (88)	4 (67)	84 (64)	618 (69)
Total	688 (77)	13 (1)	16 (2)	3 (0.3)	21 (2)	10 (1)	8 (0.9)	6 (0.7)	132 (15)	897 (100)

Table 8.24: Calculating the denominator (excluding patients dying in hospital and those discharging themselves or being discharged by a relative)

	White British	Irish	Other White	All Mixed	Indian	Other Asian	All Black	Other Ethnic	Unknown ethnicity	Total
Denominator	3082 (75)	71 (2)	79 (2)	9 (0.2)	103 (3)	34 (0.8)	45 (1.1)	28 (0.7)	625 (15)	4076 (100)
HES 2012	3205	73	81	10	105	35	46	29	629	4213
Died in hospital	117	2	2	1	2	1	1	1	3	130
Discharged by Self/relative	6	0	0	0	0	0	0	0	1	7

Numbers of cancer patients from ethnic minority groups attending the selected hospital were small as were the numbers of potential participants identified by the CRN. However, there were a large number of potential participants with unknown ethnicity which hinders interpretation. Fluctuations in proportions of ethnic minority groups can be explained by the small numbers which are not substantial enough to draw any conclusions from. Potential participants identified by ethnic groups ranged from 18% to 33% (see Table 8.25).

Table 8.26 reveals a large proportion of patients labelled as 'unaccounted' (78% in total). These patients received cancer treatment at the selected hospital but were not considered for trials. This may have been because there were no trials open to recruitment for their condition at the time of their diagnosis or they did not meet eligibility criteria. This information may be recorded in MDT data logs to which I did not have access. However, the proportions of unaccounted for patients did not differ by ethnicity. The White British and Irish groups had the lowest proportion of patients declining to participate in a clinical trial (3%) compared to 9% in the Other Asian group. Other Asians were the highest recruited (18%) compared to the lowest recruited Mixed group (11%). However, given the small sample, these findings must be treated with caution and should be subjected to further investigation in the future once more data are available.

All in all, based upon the short period of time for which trial recruitment data were available and the incompleteness of ethnicity across all datasets, it was not possible to assess the under-representation of ethnic minorities into RCTs. Improving the

collection of ethnicity data in healthcare generally as well as establishing routine collection of ethnicity data within trials themselves as well as research network trial recruitment databases, e.g. CRN recruitment data as will be demonstrated in the next chapter, would be required to answer this question in the future.

Table 8.25: Representation of potential participants identified by the Cancer Research Network of patients from the large local teaching hospital (% of denominator)

	White British	Irish	Other White	All Mixed	Indian	Other Asian	All Black	Other Ethnic	Unknown Ethnicity	Total
Potential participants identified	688 (22)	13 (18)	16 (20)	3 (33)	21 (20)	10 (29)	8 (18)	6 (21)	132 (21)	897 (22)
Denominator	3082	71	79	9	103	34	45	28	625	4076

Table 8.26: Representation of potential participants from all patients attending the local teaching hospital with a cancer diagnosis by trial status

	White British	Irish	Other White	All Mixed	Indian	Other Asian	All Black	Other Ethnic	Unknown Ethnicity	Total
Clinician decision	47 (2)	-	3 (4)	1 (11)	2 (2)	1 (3)	1 (2)	-	8 (1)	63 (2)
Patient declined	105 (3)	2 (3)		1 (11)	4 (4)	3 (9)	-	2 (7)	32 (5)	149 (4)
Being screened	59 (2)	-	-	-	-	-	-	-	8 (1)	67 (2)
Recruited	477 (15)	11 (15)	13 (16)	1 (11)	15 (15)	6 (18)	7 (16)	4 (14)	84 (13)	618 (15)
Unaccounted	2394 (78)	58 (82)	63 (80)	6 (67)	82 (80)	24 (71)	14 (31)	22 (79)	492 (79)	3177 (78)
Denominator	3082	71	79	9	103	34	45	28	625	4076

8.4 Discussion

As reported in the results section, City X has a higher proportion of ethnic minorities compared to the national average. However, data shows low proportions of these patients being treated for cancer at the local hospital in the past six years and consequently few cases being identified as potential trial participants by the cancer research network. This was not entirely unexpected given the younger age structure of ethnic minority groups in the local area. The data showed no differences by ethnicity in the proportions of patients being admitted for surgery, or dying whilst in hospital or self-discharging which may partly explain the lack of ethnic minority patients being identified as potential clinical trial participants by the research team.

There could be a number of explanations for the small number of malignancies observed in the ethnic minority groups, one of them being ethnic minorities are less susceptible to certain cancers, such as the lower rates of breast cancer in South Asians (Farooq et al, 2005). The data confirmed existing evidence of higher incidences of prostate cancers in Caribbean males (National Cancer Intelligence Network, 2009) and indicated that this increased risk may also be present in the Mixed population where one of the ethnic groups is Caribbean, African or of Other Black background.

A closer examination of the ethnic minority population structure by age and sex may shed further light onto the under-presentation problem. It is reported that

three-quarters of all newly diagnosed cancers occur in people aged 60 or over (Macmillan, 2012). However, as age and sex data by ethnic group for the 2011 census were not available at the time of this analysis, I could not assess this in the population as a whole. However, age and sex data by ethnic group at a local authority level was available in the table commissioned of the 2001 census which was used as a surrogate. I identified the population aged 50 and above in 2001 who would have been 60 years plus by 2011. The downside to using these data was that it would not include movements in and out of the West Midlands or births and deaths occurring during the period between the surveys.

The population pyramids below show the difference in the White West Midlands population to the non-White ethnic minority groups (Figure 8.2). The smaller apexes of the ethnic minority populations compared to the White population are striking but not unexpected. The shapes of the pyramids are consistent with that of the UK presented earlier in Chapter Three.

The shape of the pyramid for the White group is typical of an established population with slow growth (Figure 8.2, b). Longer life expectancy is highlighted by the broad bands in the top tertile. The ethnic minorities have a much more pyramid like shape with a broad base indicating a larger proportion of younger people and children narrowing to a small apex exposing a much smaller older population (Figure 8.2c-l).

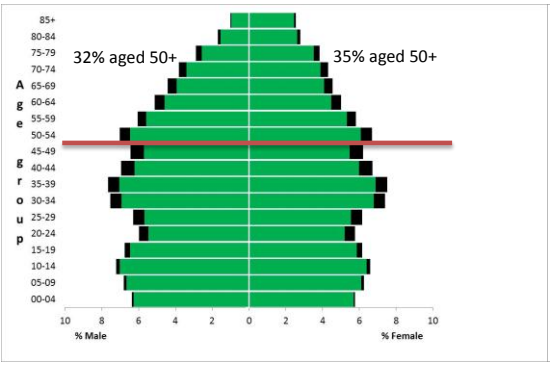
Looking in more detail at the subset with the highest risk of cancer by ethnic group (>=60 or 50 years as it is here) this proportion was 36% in the White population (Figure 8.2b). In contrast, the Caribbean group who had the largest elder population of all the ethnic minority groups comprised 26% in this age group (Figure 8.2g). The Indian group had the second largest older ethnic minority population with 19% in this group (Figure 8.2c). As these groups were earlier migrants this was not unexpected. Later migrants from the Indian subcontinent were much younger, e.g. only 11% and 10% of Pakistanis and Bangladeshis being >=60years (Figure 8.2d and e, migration patterns were discussed earlier in Chapter One). The youngest population by far was the mixed group with only 5% aged 60 years plus (Figure 8.2l).

The Caribbean and Other Black background groups experienced a baby boom between the 1950s-1970s which means in the next ten years the over 60s population in this group will increase significantly (Figure 8.2g and i). This bulge in the 40-50 year olds in the Caribbean and other Black background groups may also result in an increase in prostate cancers in this more susceptible group in the near future. The African group also displays a similar bulge but the majority were born outside of the UK suggesting a period of mass migration (Figure 8.2h). This information could help policy makers plan future resources and drive research in this area.

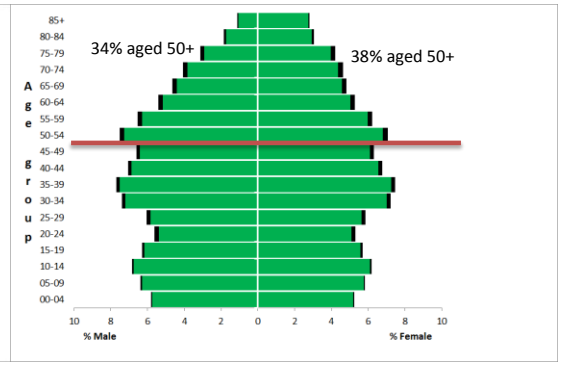
Figure 8.2: Age and sex populations pyramids by ethnicity for the West Midlands using data from the 2001 census

(darker shading indicates births outside of the UK, line across the chart indicates the population aged 50 years plus and those under)

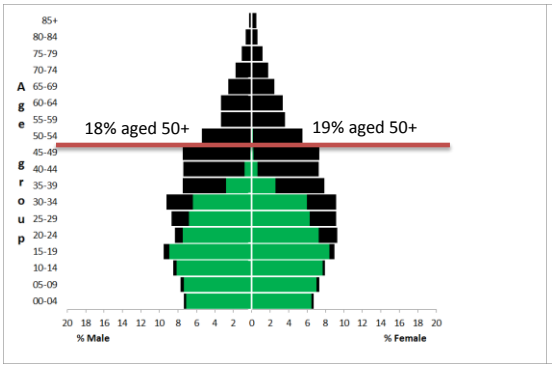
a) All (total aged 50+=34%)



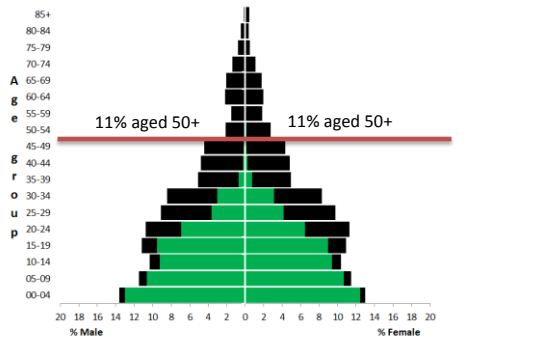
b) White (total aged 50+=36%)



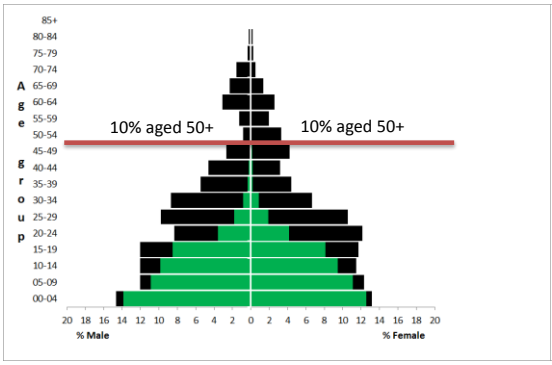
c) Indian (total aged 50+=19%)



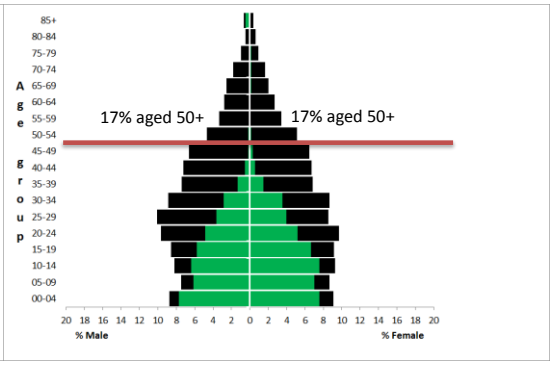
d) Pakistani (total aged 50+=11%)



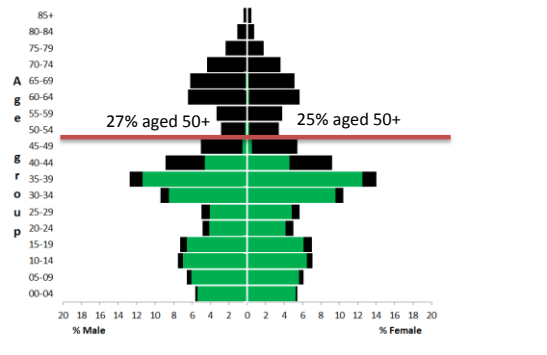
e) Bangladeshi (total aged 50+=10%)



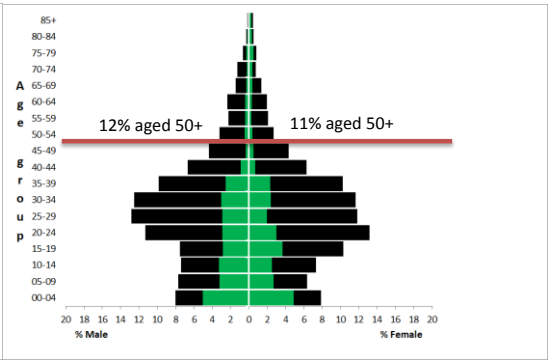
f) Other Asian background (total aged 50+=17%)



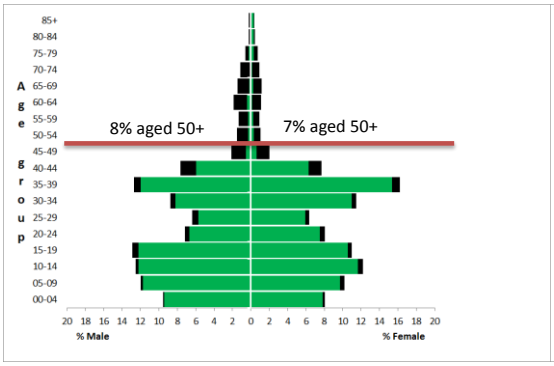
g) Caribbean (total aged 50+=26%)



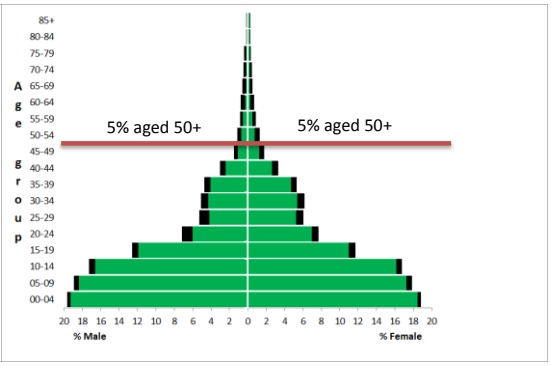
h) African (total aged 50+=11%)



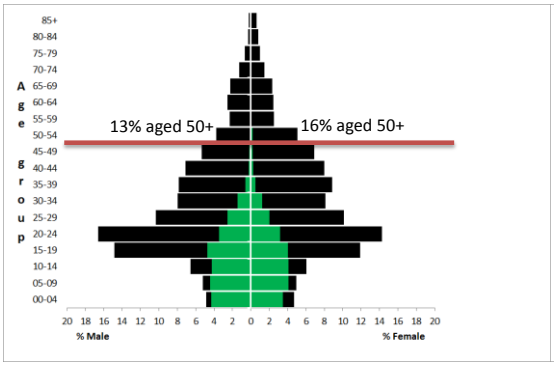
i) Other Black background (total aged 50+=7%)



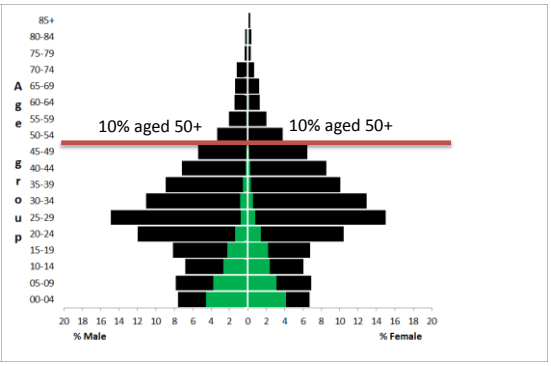
j) Mixed (total aged 50+=5%)



k) Chinese (total aged 50+=15%)



l) Other ethnic group (total aged 50+=10%)



A lack of awareness of symptoms could lead to later presentation and delays in treatment. Ethnic minorities and people from low socio-economic backgrounds were found to have a lower awareness of cancer warning signs in a study of 2208 British participants (of which 144 (6.5%) were of non-White ethnicity) (Robb et al, 2009). The most commonly cited barriers to seeking medical treatment were: 1) difficulty getting an appointment, 2) worry about wasting the doctors time and 3) fear of what might be found. The authors recommended a campaign to raise awareness of cancer warning signs. The 'Ethnic Minority Cancer Awareness Week' has being seeking to do just that by hosting events around the UK to raise awareness of risk factors, cancer symptoms and to promote the importance of screening (EMCAW, 2008).

Language barriers may have contributed to the small number of cancers observed in ethnic minority populations. Data from the 2011 census showed 49.8 million (92.3%) of the population aged three years and over reporting English as their main language. The remaining 4.2million (7.7%) reported a language other than English as their main language, the most common of these was Polish. When asked about their ability to speak English 726,000 people (1.3%) reported not being able to speak English 'well' and a further 138,000 (0.3%) reported no ability to speak English at all (Office for National Statistics, 2011). Ethnic minorities not proficient in English and not being able to communicate their symptoms fully may experience delay in diagnosis and/or referral to hospital.

Other explanations could include ethnic minorities migrating back to their countries of origin for treatment. Medical pluralism (where a community consults more than one type of therapy or medical system) is not uncommon in certain ethnic minority groups. For example, there is widespread use of Chinese medicine in conjunction with western medicine in the Chinese community (Green et al, 2006). There may also be cultural issues surrounding perceptions of disease and treatment or faith rooted perceptions (e.g. adopting a fatalistic attitude) (Austin et al, 2002; Dein, 2002; Helman, 2007). However, there is a paucity of good quality evidence from the UK to support this.

8.4.1 Limitations

The work reported in this chapter was limited by the lack of data and the difficulties in obtaining a denominator in order to assess whether under-representation of ethnic minorities in RCTs was a problem. The data obtained from HES contained a large proportion of missing and not specified/not known codes. Similarly, trials recruitment data from the Cancer Research Network was available for a one year period only and also contained a large amount of missing data. Many patients were present in the HES data having been diagnosed with cancer but were not accounted for in either the surgery, discharge or Cancer research network data which held information on all potential participants. It is not known why these patients were not identified but it may have been largely due to a lack of trial availability or ineligibility. Accessing data from multi-disciplinary team meetings could shed light

on this issue and allow us to gauge the proportions of patients for whom no trials are available.

At the time of writing, only preliminary data from the 2011 census had been released, data by age and gender by ethnic group were not available.

8.5 Conclusion

Population structures of ethnic minority groups are without a doubt rapidly changing within the UK. The 1991 census reported over 3 million people (5.5%) of the UK population identifying themselves as belonging to a non-white ethnic group. By 2001, this had increased to 4.6 million (7.9%) and in 2011 this increased again to 14.1% in England and Wales (UK figure not yet available) (Office for National Statistics, 2011). These changes could have been influenced by a number of factors, such as new migrants, second and third generation UK born offspring of the original migrants, as well as the numbers of people self-reporting as belonging to an ethnic minority group increasing, possibly due to the availability of more sensitive categorisations, e.g. introduction of Mixed, Arab, Traveller categories and the ability to write-in responses.

Despite the increasing proportion of ethnic minorities, I was not able to confirm the under-representation of these groups in clinical trials at a local level. However, the under-representation of ethnic minority groups in RCTs continues to be observed in ad hoc studies, such as that reported by Godden et al (2010) who found, when comparing patients admitted with cancer after adjusting for age, disease and

gender the odds of being in a trial 30% lower in ethnic minority cancer patients compared to White cancer patients. However, there is evidence to suggest ethnic minorities are better represented in trials of other disease areas such as heart disease (Jolly et al, 2005, Cooper et al, 2003). Jolly et al (2005) reported exclusion rates and reasons for exclusion in a study of cardiac rehabilitation conducted in ethnic-rich Birmingham. Results revealed patients of South Asian background to be the most likely to be excluded and for the exclusion to be language based. Reassuringly, recruitment of eligible patients was similar across ethnic groups with South Asians accounting for 17% of the final sample. The external validity of RCTs becomes questionable when certain groups have been excluded from a trial (Rothwell, 2005). Ethnic minority patients need to be represented in RCTs in order to ensure generalisability to the wider UK population.

The biggest challenge in assessing under-representation remains the incompleteness and questionable quality of ethnicity data. Ethnicity data must continue to be prospectively collected by Cancer registries across the UK in order to obtain a more accurate picture of cancer incidence. Ethnicity in HES databases also requires improvement in both inpatient and outpatient records. The current data contains large numbers of unknown ethnicities. Some suggestions as to how this could be remedied were made earlier in the literature review of improving ethnicity data collection (see Chapter Four). Lastly, the importance of recording ethnicity at research network level as well as in RCTs themselves needs to be recognised in order to assess under-representation accurately in the future.

8.6 Summary

In this chapter, I assessed the completion of ethnicity data in local HES datasets. I explored the population structure of ethnic minorities on a national level and for the West Midlands and identified proportions >=60 who are most at risk of developing cancer. I also scrutinised the Cancer Research Network recruitment data and used this information to assess representativeness of ethnic minority patients entering cancer trials.

In the next chapter, I describe my journey towards establishing ethnicity data collection at a local CRN and describe the problems I encounter along the way. I then attempt to use the collected ethnicity data to assess the representation of ethnic minorities in clinical trials and explore barriers to clinical trial participation by ethnicity using data from the selected Cancer Research Network recruitment database.

Chapter 9 Ethnicity data collection and participation in clinical trials

9.1 Introduction

As discussed in Chapters One and Seven, it has been reported that ethnic minorities are under-represented in clinical trials. This is based on the known size of the ethnic minority population and the reported incidence of disease in that population. (Mason et al, 2003, Hussain-Gambles, 2004 and Goddard et al, 2012). Two systematic reviews carried out prior to the formation of the NCRN suggested that clinicians may be reluctant to approach ethnic minority patients for inclusion in clinical trials due to time restrictions and a lack of resources (Ross et al, 1999, Ward et al, 2000). Ward et al (2000) and Hussain-Gambles et al (2004) reported the problem of under-representation stemming from a fear of authority and lack of faith in the medical profession.

The Equality Act 2010 holds organisations such as the Health Service responsible for ensuring there is no discrimination in access to health care on the grounds of any of its nine protected characteristics, one of which is race (Home Office, 2010). As patients who enter clinical trials are reported to have better outcomes, there is therefore an urgent need for evidence on barriers to recruitment into RCTs by ethnic group before strategies can be developed to address these barriers and improve uptake into trials (Braunholtz et al, 2001). Ethnicity data collection within

the context of the clinical trial recruitment network is also needed in order to address this problem.

One of the problems which I encountered and only partially solved was the reluctance of healthcare professionals to collect ethnicity data. Focus groups with healthy participants reported in Chapter Five revealed ethnic minority groups, more specifically South Asians, were willing to provide their ethnicity data for healthcare purposes, provided they are given an explanation as to why the data are being collected and how it will be used to improve services.

Part of my research included establishing ethnicity data collection at a local Cancer Research Network (CRN). The selected CRN was established in 2001 and is committed to entering patients into trials. By 2006, it had exceeded the NCRN recruitment target of entering 7.5% of cancer patients into clinical trials by successfully recruiting 35% of all cancer patients treated through their network (NCRN, 2006). The CRN routinely collects trial recruitment data for monitoring and reporting purposes as part of their recruitment strategy. Once a patient has been identified as a potential participant, they are screened by a member of the research network team operating in that particular disease site and offered a trial if suitable. The reason why a patient may be considered unsuitable by the clinician is recorded in a recruitment database. In circumstances where eligible patients choose not to participate in a trial, the reasons they give for deciding so are recorded in an independent anonymised 'Reasons for refusal' database.

This chapter provides a narrative of my journey towards establishing ethnicity data collection at a local cancer research network and was broken down into two sections; the first being the establishment of ethnicity data collection as part of the 'reasons for refusal' dataset at the selected CRN. In the second part, I attempt to use the data collected in part one to investigate the representation of ethnic minorities in clinical trials and reasons for non-participation by ethnic group.

9.2 Timeline of establishing ethnicity data collection

In order to achieve ethnicity data collection on all patients identified by the CRN as potential participants, I proposed obtaining prospective self-reported data from patients. The process of my efforts to supplement the 'reasons for refusals' database with ethnicity is summarised as milestones in chronological order below and is described in further detail in each of the relevant sections:

- Jan 2008 - I had an initial meeting with the cancer research network lead and managers
- Oct 2008 - I identified items of interest and developed ethnicity questionnaire
- Dec 2008-I finalised ethnicity questionnaire with network lead and managers, received approval to collect ethnicity data from the Caldicott Guardians
- Apr 2009 - I met with Comprehensive Local Research Network (CLRN) to discuss ethnicity data collection through their network

- May 2009 - I received confirmation that ethics approval was not required to add ethnicity questions to existing data collection from the local ethics committee chair who the deemed the exercise to be service evaluation
- Jun 2009 - I met with database administrative staff (clinical trials officers) to discuss practicalities of adding ethnicity fields to the existing database to capture data from completed questionnaire and data entry
- Jul 2009 - I met with clinical trials officers, network manager & research nurses to discuss the questionnaire and point of administration. Concerns were raised resulting in the proposal to collect prospective self-reported ethnicity data via nurse administered questionnaire being rejected.
- Jul 2009 –Questionnaire adapted to be suitable to be posting to patients
- Jan 2010 - Obtained relevant approval from the selected hospital and its peripheral sites
- Feb 2010 – I provided first batch of 40 questionnaire packs to CRN
- Mar 2010 - 100 further packs provided
- Apr 2010 – A questionnaire was sent out to non-declining patient in error
- Feb 2012 - Exercise terminated due to upcoming change of CRN database
- Apr 2012 - Change of CRN software. Staff now have ability to record ethnicity data and reasons for non-participation recorded directly on database

9.3 Proposed method 1: Ethnicity data collection by research nurses at point of refusal

Once a patient has declined to participate in a trial they are routinely asked by the CRN if they would be willing to provide a reason for their decision. I proposed that ethnicity data were requested at this time through a short nurse administered questionnaire. The work was deemed service evaluation so did not require ethics approval, and as no patient identifiers were included in the questionnaire I did not require Ethics and Confidentiality Committee (ECC) approval.

I developed the ethnicity questionnaire in conjunction with the CRN, it comprised of six ethnicity questions, 1) ethnic group, 2) religion, 3) country of birth, 4) first language or mother tongue, 5) language in which the patient feels most comfortable speaking with healthcare professionals and 6) ability to speak and understand English (see Appendix 3). Responses to questions on ethnic group, religion and country of birth were categorised tick boxes with space provided for the entry of free text should the categories be inadequate. In addition, where country of birth was outside of the UK or Republic of Ireland the number of years residing in the UK were requested. Data on first language and the choice of language a patient feels most comfortable speaking with their doctor or nurse captured free text only. Patients were also asked to rate their ability to speak English using a Likert scale with responses available of Excellent, Good, Fair, Poor and Not at all.

The ethnicity questions and offered response categories were based upon the findings of the ethnicity data collection literature review reported in Chapter Four. Both the questions and response categories conformed to the Department of Health (2005) guidelines for ethnicity data collection in the NHS. Ethnic group was collected using the 16+1 categories which map directly onto the 2001 census data. A question on the reason for declining a trial with a free text response was included at a later date (details below).

9.3.1 How the proposal was received

I had an initial meeting with the CRN lead and network manager to discuss incorporating ethnicity data collection to enhance the existing anonymised 'reasons for refusal' database. The idea was well received and of great interest to the network lead. The CRN spans across five NHS trusts with the selected hospital hosting the main site. Authorisation from the research and development managers was needed for each of the four peripheral hospital sites. I gave an initiation presentation explaining the problem of ethnic minority under-representation and the need for the collection of ethnicity data in conjunction with reasons for refusal to the research team prior to commencement.

I discussed the practical issues of data collection, primarily data entry and database issues with the administrative staff or Clinical Trials Officers who would be responsible for processing the data. Both Clinical Trials Officers were reluctant to add items to the database and suggested a separate data file for ethnicity. This was

not feasible as the database that held the reasons for refusal was anonymised, and ethnicity would need to be stored in the same database in order to maintain the link between the two data items. The Clinical Trials Officers also expressed concerns about how this would fit into their already busy work schedule. Their concerns were discussed with the Network Manager and both issues were addressed through agreement to add a 'reason for refusal' question to the questionnaire and I volunteered to take on the task of data entry.

The ethnicity and reason for refusal questionnaire was reviewed by research nurses and clinical trials officers and I met with them to obtain feedback. Concerns relating to an ethnicity question and the administration were voiced. One individual strongly disliked the idea of asking patients how long they have lived in the UK (where country of birth was non-UK/ROI). The individual strongly felt this would alienate patients and invoke feelings of being singled out, particularly if they were recent immigrants. The individual raising the concern was a non-UK born resident and spoke from personal experience. No other objections to the individual questions were raised. However, strong feelings were expressed by two research nurses regarding the administration of the questionnaire. They felt that asking for ethnicity data would be too much for patients who were already overwhelmed and vulnerable. Most would be trying to come to terms with a recent cancer diagnosis or relapse. Patients would also be trying to digest possible treatment options and/or trial options or, at the very least, given study material to take away and read/digest.

The Research Nurses spoke of patients' struggle to process study information and the frequency with which they were confronted with blank looks and knew the information had not been retained. A proportion of patients rejected the idea of going into a study immediately and therefore were never approached by the research network. Others were approached, but did not wish to go into a study or even take the information home to read and consider later. In this situation the patient is not asked any further questions, including reason for refusal.

The nurses also spoke of the guilt expressed by patients who, having spent time with the consultant or research nurse felt as though they were letting them down by declining to participate. Patients were overly apologetic and sometimes asked if there was anything else they could do to help. However, the nurses on seeing the patients were feeling needlessly guilty and not wishing to put them through anymore anxiety, felt asking ethnicity at this point would be taking advantage of their fragility.

All in all, the nurses did not feel the timing of the questionnaire was appropriate nor felt comfortable asking the patients for this information verbally. Obtaining such data from patients using a research nurse administered questionnaire was rejected.

9.4 Proposed method 2: The postal questionnaire

Research nurses at the CRN rejected my proposal to administer the ethnicity questionnaire at the point of refusal, but believed posting the ethnicity

questionnaire to patients declining the offer to participate in a clinical trial to be more acceptable. Cancer patients of all ethnic groups (including white) attending the CRN and who were eligible to enter a clinical trial and declined were mailed a questionnaire post trial refusal.

The ethnicity questionnaire I developed was amended to include basic patient information, such as age group, gender, and type of cancer and name of the trial which they were offered. All patients were encouraged to return the questionnaire with an additional tick box provided for patients' not wishing to complete the information included as the first item appearing at the top of the questionnaire. Questionnaires were mailed to patients by research nurses with a covering letter from the nurse who originally discussed the trial with the patient. The pack contained a copy of the questionnaire, the cover letter introducing the questionnaire, the need for it and how it will be used, and a pre-addressed postage paid envelope. In total, 160 packs were supplied to the CRN.

9.4.1 Implementation of the postal questionnaire exercise

Questionnaires were posted to cancer patients declining to participate in clinical trials by the research nurses. It was assumed the accompanying cover letter from the nurse who had personal contact with the patient would have a positive effect on the response rate.

I was able to keep tally of the numbers of questionnaires sent out through regular reports from the CRN and compare this with the numbers of completed forms

returned to me in the pre-addressed envelopes and therefore calculate a response rate.

After receiving 11 completed questionnaires, I was contacted by telephone and letter by a patient (referred to as Mrs X from this point onward) regarding a questionnaire she had been posted. Mrs X had not refused to participate in a trial and, in fact, was already participating in a clinical trial when she received information about a second trial. Prior to making a decision on whether to participate in the second trial she received a phone call from the network apologising as a mistake had been made and she wasn't in fact eligible for the second trial. She subsequently received one of my questionnaires. Mrs X did not understand why she had been sent the questionnaire and felt she has been labelled negatively as a 'refuser' which as a participant in another trial she was not. Mrs X disapproved of the term 'refusal' stating that the choice of word 'refusal' was unfortunate as patients are neither asked nor told to enter a trial. She explained patients are given the opportunity to participate in a trial. It is explained to them; they take the information away, weigh the pros and cons and then make a choice. Mrs X pointed out that, had she been eligible for the second trial, she would not have participated after her experiences with the first clinical trial in which she was participating. This was mainly due to the burden of extra visits to the hospital which were time-consuming and strenuous.

The research network was alerted and the incident investigated. It was raised at team meetings to ensure questionnaires were only sent out once a patient had declined to participate.

A total of 13 nurses took part in the postal questionnaire survey, with one taking maternity leave during the course of the exercise. Three of the remaining 12 nurses, posted out questionnaires to declining cancer patients (25%). In total, 16 questionnaires were mailed out to patients. Ten completed questionnaires were received in total and a further one where the patient did not agree to complete the form, resulting in a response rate of 69%. No further questionnaires were received after the incident with Mrs X.

Feedback from the CRN manager suggested the main reason for the low rate of questionnaires being posted out was because the nurses forgot as it was not part of their usual routine. However, some nurses felt that once a patient refused to participate in a clinical trial, that no other surveys or studies should be offered or entered into. The postal questionnaire exercise was terminated after a running for a period of two years (Feb 2010 to Feb 2012). Although severely limited, data collect from the completed ethnicity and reason for refusal questionnaires are presented below.

9.4.2 Postal questionnaire results

Ten patients of the 16 who were posted a questionnaire returned a completed form in the two year timeframe. One further patient returned an incomplete

questionnaire having ticked the box to say he/she was not happy to provide the requested information, and one non-declining patient was posted a questionnaire in error.

Respondents were balanced across the age groups and half did not provide gender (n=5). All bar one were of White ethnicity, including White British (6), White Irish (2) and White Other (1). The majority were of Christian faith, were born in England and spoke fluent English.

Eight patients suffered with cancers of the head and neck, the remaining two had ovarian and prostate cancer (Table 9.1). The reasons for refusing a trial were varied. Five themes were identified (the patient ID to which it relates is listed in brackets):

- Treatment preference (1, 3, 5, 10)
- Patient does not wish to be randomised (2)
- Patient has been through enough/already stressed (4, 9)
- Burden of extra hospital visits (6, 8)
- Burden of trial related activity (7)

The most common reason for non-participation in this small set was the preference for a particular treatment. Eight patients were diagnosed with head and neck cancers and were offered the PET-NECK trial (based at Warwick). This trial aimed to compare the efficacy of a PET-CT guided watch and wait policy with the current practice of planned neck dissection. Patients were opposed to the option of watching and waiting and therefore chose not to be randomised.

Three patients rejected a trial based upon the number of extra hospital visits involved. One patient with a needle phobia rejected the trial as it would involve injections/blood draws. One patient felt they had been through enough already and another was too stressed following diagnosis to consider a trial.

Table 9.1 Ethnicity, cancer type, and reasons for refusal per patient

ID	Ethnicity	Type of cancer	Reason for refusal
1	White British	Tongue and throat	I wanted surgery on my lymph glands before radio/chemo therapy and the trial would have randomised this decision.
2	White British	Tongue	It was recommended by the oncologist that I go straight to radio/chemo therapies. The trial was random and surgery first would have delayed treatment.
3	White other	Prostate	I prefer to have regular blood tests and if my PSA increase then I will have radiotherapy
4	White British	Left paratoid gland	I had concerns that it may not be as effective as the original radiotherapy treatment I along with my family researched on the internet and also consulted an independent doctor for advice all of which made up my mind that it was too much of a risk. Having gone through an operation lasting over 6 hours I feel it is time to think of myself and my family after having a traumatic few months.
5	White British	Paratoid	Risk of all areas not being treated by radiotherapy
6	White Irish	Thyroid	Distance and time to Coventry 25 miles each way
7	White British	Thyroid	I am extremely needle phobic and don't want to have any more injections than necessary. That is sole reason.
8	White Irish	Ovarian	Too many extra trips to hospital which would be stressful
9	Asian Indian	Thyroid	Already stressed about my condition
10	White British	Throat	I have gone for surgery straight away then Chemo therapy /radio therapy. You can use any of my notes for research.

9.5 Proposed method 3: Transfer of ethnicity data from other sources

Old software at the selected CRN which prohibited the collection of ethnicity data was phased out in April 2012 and replaced with new database (EDGE) which included a dedicated field for ethnicity and a coded list for 'reasons for non-participation'. This presented me with a new opportunity to initiate the prospective collection of ethnicity data. The CRN proposed transferring ethnicity data from Hospital Episode Statistics (HES) or medical notes (where recorded) into the new database. The exercise commenced in April 2012 at which time the new database was activated.

The second section of this chapter involved utilising the collected ethnicity data to assess the representation of ethnic minorities recruited into clinical trials across the selected CRN and explore barriers to participation by ethnic group.

Once a patient has been identified by the CRN as a potential participant, they either fail to meet the eligibility criteria and are not offered the trial and the 'recruitment status' field on the database is coded as 'clinician decision', decline to participate which is recorded as 'patient decision' or are 'recruited'. Patients undergoing tests for eligibility are recorded as 'screening'. Trial status by ethnic group is presented in this chapter. In addition, reasons given by patients for choosing not to participate in a trial and the reasons why clinicians deemed them unsuitable is explored with and without ethnic group.

Reasons for non-participation were analysed using a qualitative thematic approach. Responses were coded and categorised into themes and subthemes. Analysis in the first instance was blind to ethnic group. Further analysis by ethnicity was performed once themes had been established. Ethnic groups were combined due to small numbers into Whites and non-Whites.

9.5.1 Results from method 3

Data on a total of 1271 potential participants were collected from April 2012 to April 2013 (Table 9.2). Three of the 1271 patients had missing recruitment data. Reasons for non-participation were collected for 331 patients who did not enter a clinical trial with 283 of these being patient decision and the remaining 48 clinician decision.

Table 9.2 Trial status for potential participants identified by the selected Cancer Research Network across all sites

Recruitment status	White British	Irish	Other White	All Mixed	Indian	Other Asian	Caribbean	Other Black	Other Ethnic	Unknown Ethnicity	Total
Clinician decision	38 (4)	1 (7)	-	1 (33)	-	-	-	1 (17)	-	7 (3)	48 (4)
Patient declined	204 (22)	2 (13)	7 (28)	1 (33)	6 (29)	5 (42)	-	-	2 (33)	56 (24)	283 (22)
Screened*	96 (10)	-	-	-	-	-	-	-	-	20 (9)	116 (9)
Recruited	605 (64)	12 (80)	18 (72)	1 (33)	15 (71)	7 (58)	2 (100)	5 (83)	4 (67)	152 (65)	821 (65)
Total	943 (74)	15 (1.2)	25 (2)	3 (0.2)	21 (2)	12 (1)	2 (0.2)	6 (0.5)	6 (0.5)	235 (19)	1268 (100)

Note * ='Screened' represents patients who were being screened for eligibility at the time of data lock

Reasons for non-participation were selected from a drop down list (predefined by database developers) and did not allow the entry of free text. The reasons given were categorised into themes based upon whether it was the patients or the clinician's decision not to participate, or other reason. A large proportion of patients (n=114, 34%) did not participate in trials but did not provide a reason or were not asked to provide a reason, leaving 217 patients with data. Patient decision accounted for 47.5% of non-participants with reason for non-participation data, clinician decision for 38.7% and other reasons 13.8% (Table 9.3). Patients not put forward for trials by the clinicians were mainly due to ineligibility (31.8%). Clinicians decided 3.7% of patients had been through enough, a further 3.2% were coded as 'inappropriate for approach', but no further information was available. The most common 'other reason' for non-participation simply stated 'Other' indicating a suitable match from the drop down list had not been found. Two patients died between being identified and approached and one patient was already participating in one trial, but it is not clear who made the decision not to participate in the second trial. It may have been the patient not wishing to partake in another trial or clinician decision if the two trials were incompatible.

Table 9.3 Reasons for non-participation

Reason for non-participation	All patients		Patients with reasons data	
	N=331	%	N=217	%
No reason reported	114	34.2	-	-
Patient decision	103	31.2	103	47.5
Cannot cope with thinking about a trial and making another decision	39	11.8	39	18.0
Does not want extra visits to hospital	10	3.0	10	4.6
Does not want extra blood tests	2	0.6	2	0.9
Inconvenience of trial treatment	5	1.5	5	2.3
Treatment too long	1	0.3	1	0.5
Put off by toxicities	10	3.0	10	4.6
Patient has treatment preference	7	2.1	7	3.2
Patient doesn't wish to be randomised	6	1.8	6	2.8
Wants the gold standard treatment & doesn't want to be a guinea pig	13	3.9	13	6.0
Wants treatment elsewhere	4	1.2	4	1.8
Not satisfied with care services	1	0.3	1	0.5
Patient doesn't want treatment	5	1.5	5	2.3
Clinician decision	84	25.5	84	38.7
Clinical decision patient has been through enough	8	2.4	8	3.7
Inappropriate for approach	7	2.1	7	3.2
Ineligible	69	20.9	69	31.8
Other reason	30	9.1	30	13.8
Death	2	0.6	2	0.9
Already participating in another trial	1	0.3	1	0.5
Other reason, not specified	27	8.2	27	12.4

The most common reason given by patients choosing not to participate was the inability to cope with thinking about a trial or being faced with another decision (18%). Other patients refused on account of the burden of extra visits to the

hospital (4.6%) or extra blood draws (0.9%), one patient stated that the treatment period was too long (0.5%). Five patients declined because of the inconvenience of the trial treatment (2.3%). As all of the above related to the inconvenience aspect of the trial, they were grouped together to form a subtheme which accounted 8.3% of patient decisions. Further subthemes were identified by grouping similar reasons as follows (also see Table 9.4):

1. Feeling overwhelmed
2. Inconvenience
3. Toxicity
4. Treatment preference
5. Does not wish to be experimented upon/randomised
6. Changing hospital
7. No treatment

Patients 'not wanting to be experimented on' or 'randomised' was the second most common theme accounting for 8.8%. The list of treatment toxicities put off a 4.6% of patients, with a further 3.2% having a treatment preference, and 2.3% declining all treatment. Changing of hospital accounted for 2.4% with one patient reporting this to be due to unsatisfactory care services.

Table 9.4 Patient decision subthemes and their components

Subtheme	N=103	%
1. Feeling overwhelmed	39	18.1
Cannot cope with thinking about a trial and making another decision	39	18.1
2. Inconvenience of trial	18	8.3
Does not want extra visits to hospital	10	4.6
Does not want extra blood tests	2	0.9
Inconvenience of trial treatment	5	2.3
Treatment too long	1	0.5
3. Treatment toxicity	10	4.6
Put off by toxicities	10	4.6
4. Treatment preference	7	3.2
Patient has treatment preference	7	3.2
5. Does not wish to be experimented upon/randomised	19	8.8
Patient doesn't wish to be randomised	6	2.8
Wants the gold standard treatment & doesn't want to be a guinea pig	13	6.0
6. Changing hospital	5	2.4
Wants treatment elsewhere	4	1.9
Not satisfied with care services	1	0.5
7. No treatment	5	2.3
Patient doesn't want treatment	5	2.3

The data were weak, with the majority of cases being of White British origin and unknown ethnicity and with insufficient numbers from ethnic minority groups to derive any conclusions.

The most common reason for clinicians not offering a patient a trial was ineligibility (Table 9.5). However, seven patients (6 of White British or Irish origin and one with unknown ethnicity) were considered inappropriate for approach.

Table 9.5 Clinician decision or other reason for non-participation by combined ethnic group

Clinician decision	White	Non-White	Unknown	Total
Patient has been through enough	7		1	8
Inappropriate for approach	6		1	7
Ineligible	54	5	10	69
Subtotal	67	5	12	84
Other reasons				
Death	2			2
Already participating in another trial			1	1
Other reason, not specified	15		12	27
Subtotal	**17**	**0**	**13**	**30**

Investigating patient decision by ethnic group revealed the most common reason provided for the non-participation of White patients was found to be the feeling of being overwhelmed and not being able to think about a trial or make a decision (Table 9.6). Although, seven patients with unknown ethnicity gave the same reason, no non-White ethnic minority patients in this extremely small sample offered this reason. Three patients from ethnic minorities had a treatment preference, three did not wish to be randomised or experimented upon, and two patients were put off by treatment toxicities listed in the patient information leaflet.

Table 9.6 Patient decision by combined ethnic group

Patient decision themes	White	Non -White	Unknown	Total
Declined to take part, no reason reported	93	4	16	113
1. Feeling overwhelmed				39
Cannot cope with thinking about a trial and making another decision	32		7	39
2. Inconvenience of trial treatment				18
Does not want extra visits to hospital	7		3	10
Does not want extra blood tests	2		1	2
Inconvenience of trial treatment	4		1	5
Treatment too long	0			1
3. Treatment toxicity				10
Put off by toxicities	4	1	5	10
4. Treatment preference				7
Patient has treatment preference	4	3		7
5. Does not wish to be experimented upon/randomised				19
Patient doesn't wish to be randomised	5	2	1	6
Wants the gold standard treatment & doesn’t want to be a guinea pig	9		2	13
6. Changing hospital				5
Wants treatment elsewhere	4			4
Not satisfied with care services	1			1
7. No treatment				5
Patient doesn't want treatment	4		1	5
Subtotal	**169**	**10**	**37**	**216**

9.6 Limitations

I conducted a practical exercise to collect data prospectively in a set timeframe of a year with a team of trial staff who were inexperienced in ethnicity data collection. The exercise demonstrated the difficulty of collecting ethnicity data particularly due to staff perceptions and the perceived usefulness of the data, and the extent to which staff felt comfortable in collecting these data. Similar issues were apparent in the survey of healthcare professionals (Chapter Six). One problem was that the database in use did not allow the capture of ethnicity data. Changeover to a new database, which came with dedicated ethnicity and reason for non-participation fields was the most effective method in changing staff behaviour and resulted in the recording of ethnicity data.

The postal ethnicity questionnaire was unsuccessful not because patients did not complete and return the questionnaire but because very few members of staff remembered to send the questionnaire out, whilst others actively declined to send it out because they believed once a patient says no to research they should not be contacted any further. Conducting research on people who have decided not to participate in research was a difficult subject to explain and cropped up often throughout my research journey.

9.7 Summary

The biggest barrier that I encountered in attempting to collect data on ethnic minority representation in clinical trials was the research nurses reluctance to ask patients to provide ethnicity data. This confirms the findings of the ethnicity data collection literature review reported in Chapter Four and the survey of healthcare professional (HCP) perceptions of ethnicity data collection reported in Chapter Six. Fear of offending patients and a lack of understanding of the importance of these data were identified in both of these pieces of research

The research nurse administered questionnaire proposed in method 1 was rejected outright as the nurses were not comfortable asking patients for this data although there was no objection to the questions themselves except for that raised by the Clinical Trials Officer regarding the numbers of years lived in the UK for non-UK born ethnic minorities. The postal questionnaire, although proposed by the research nurses as an alternative to asking patients for these data, was unsuccessful with only 16 questionnaires sent out over a two year period. The response rate however was 73% (11 of 15 were returned excluding one sent out in error), indicating patients were generally happy to provide such information with only one patient returning the form uncompleted. It is my belief that no further questionnaires were posted out following the incident with Mrs X.

The restrictions imposed by the database used at the time of the start of this exercise did not have fields within which to record ethnicity, if collected it would

have been held on an independent datasheet. A new replacement database allowed the transfer of ethnicity data from other hospital sources to the new database, and this was the most effective solution. However, this system for collecting ethnicity data is dependent on other healthcare professionals collecting ethnicity data and does not address the barriers preventing research nurses collecting this information. In order for improvement this method requires full collection of ethnicity data and the barriers to participation. Over, 34% of patients not participating in a trial had no reason recorded.

The network manager stated time restrictions and a lack of resources to be the main reasons why ethnic minority patients are not offered a trial. Additionally, the network manger believed Multi-Disciplinary Team (MDTs) where potential trial participants are usually identified is where they were missed. The manager did not perceive patient barriers to be the main barrier to trial recruitment, instead believing the problem lay with healthcare professionals and a lack of resources. It can take a long time to get an interpreter so staff at the CRN actively encouraged patients to bring an English speaking relative to appointments, believing this to be the best option at the moment.

Identifying and addressing HCP barriers to ethnicity data collection, e.g. raising awareness of the importance of collecting these data and using the data collected in reports to demonstrate equality in access to clinical trials is needed. Improved collection of ethnicity data will enable policymakers, researchers and healthcare providers to identify vulnerable groups and allocate resources appropriately to

target barriers effecting ethnic minority access to and participation in clinical trials, e.g. help consenting patients into studies.

Chapter 10 Discussion and conclusions

The aim of this thesis was to investigate the collection of data on ethnicity and the recruitment of ethnic minorities into clinical trials and also to determine whether ethnic minorities are under-represented in trials. I have not found evidence that ethnic minorities are under-represented, but there is insufficient evidence to demonstrate adequate representation or equality in access to RCTs. Inequalities in access to clinical trials would be in direct violation of the Equality Act 2010 and the goals set by the NHS plan to tackle inequalities and provide culturally appropriate care (Home Office, 2010; Department of Health, 2011). The inclusion of ethnic minorities in trials is important for ensuring that the trials have external validity and that the results can be extrapolated to the whole British population.

The age structure of the UK ethnic minority population is changing and ageing. The Mixed group is reported to be the fastest growing (Pinnock, 2009 and Office for National Statistics, 2011). The health service needs to adapt in order to meet the needs of an ageing ethnic minority population and subsequent increase in the burden of disease amongst ethnic minorities who were largely born outside of the UK. Future generations of current immigrants will speak English and may be more aware of health research than their parents and grandparents. People may also have more ownership of the NHS especially with non-White people coming into the NHS as professionals. Although the need for interpreters and translated trial material may decrease for the established ethnic minority groups, this will not

necessarily affect non-language specific barriers to accessing clinical trials, such as lack of trust, lack of resources and institutional racism (Macpherson, 1999). In addition, patient perceptions of research, clinical trials and what it means to participate in a clinical trial as well poor ethnicity data collection need to be addressed, both in policy and in further research.

10.1 Data collection

10.1.1Recording

There needs to be better completed, validated, self-reported ethnicity data in order to estimate the representativeness of ethnic minorities in clinical trials. One of my key findings was the inadequate completion of ethnicity data and the high proportion of missing data. Data collection has improved over recent years but there is still considerable room for improvement given the large number of patients with missing codes (not stated, not known and not specified) recorded in place of ethnicity. Analyses of local HES data revealed inpatient data to be better completed than outpatient data. This may be on account of outpatients' clinics being busier and therefore reception staff having less time in which to obtain data from patients. It may also be due to the advantage of having more time in which to obtain the information or perhaps needing to be more thorough when a patient is admitted.

Providing ethnicity monitoring training and promoting awareness of the importance of ethnicity data and its uses are essential. Collecting ethnicity data in primary care and linking through to other services would be the most efficient method of

collecting these data once resources are put in place. Success stories include the achievements by Lambeth and Central Liverpool NHS Primary Care Trusts where ethnicity data were pursued with vigour (Adebayo and Mitchell, 2004; Race for Health, 2006). Lambeth Primary Care Trust introduced the '*Individual Patient Registration Profile*' programme which provided funding and training to GP practices as an incentive to collect patient profiling data. Patients who were already registered with the practice were posted a patient profiling questionnaire with free return envelopes and data entry was funded by the programme. Templates were provided to capture patient profiling data prospectively for all new registrations.

10.1.2 Use of existing ethnicity data

Ethnicity data have been used in reports by the NCIN since 2009 through use of data collated from cancer registries, HES and ONS held in the National Cancer Data Repository. Cancer incidence data by ethnicity highlighted the flaws in the data and the large proportion of missing ethnicity (NCIN, 2009). However, the data have been used in the second All Breast Cancer Report which found women of Black origin to be younger at diagnosis compared to White women (median age at diagnosis 50 and 62 years respectively) (Lawrence et al, 2011). In addition, fewer Black and Asian women of routine screening age of 50-70 years were reported to have had their cancer detected through this method compared to White women in the same age group (44.6%, 52.1% and 56.3% respectively). Further to this, 56% of Black women were reported to be more likely to have poorer grade 3 tumours and

64% had lymph node involvement compared to White Women with 36% and 38% respectively.

Ethnicity data collected by Lambeth Primary Care Trust have been used in a health equity audit of Stop Smoking Services and a needs assessment exercise of the Portuguese community (Race for Health, 2006). Ethnicity data have also been used to investigate the poor uptake of screening services in detail within ethnic minority populations which may partly explain the lower rates of cancer observed in this thesis (Chapter Eight). Szczepura et al (2008) reported uptake patterns for breast and bowel cancer screening in UK South Asians vs non-South Asians over a 15 year period, and reported an improvement in screening rates as a result of targeting single GP practices. Breast cancer screening rates were generally found to be lower in South Asians than non-South Asians (61% and 75% respectively). However, an improvement in overall breast cancer screening was observed during the course of the study to 67% in South Asians and 78% in non-South Asians. Bowel cancer screening rates were found to be significantly lower in South Asians vs non-South Asians (33% vs 61% respectively). Bowel screening rates were reported to be particularly low in the Muslim population (26%).

In a study of attitudes to colorectal cancer screening in UK ethnic minorities, Robb et al (2008) reported less knowledge of the causes of colorectal cancer in ethnic minority participants when compared to White British participants. Feelings of 'embarrassment' and 'shame' were the most cited barriers to screening in the ethnic minority communities.

10.1.3 Perceptions and experiences of ethnicity data collection

Focus group participants did not have many issues with regards to providing their data for healthcare purposes, on condition that an acceptable rationale was given. By contrast, one third of healthcare professionals did not attempt to collect any data because either they or their organisations did not believe it to be of importance. Several respondents said they did not understand why such data were required, possibly due to a lack of training. This lack of understanding could explain why some Healthcare professionals (HCP) chose not to ask the patient to provide these data, and instead opted for assessment by observation or an alternative method whereby they would not need to offer an explanation to the patient. One HCP survey respondent thought staff may not be aware they needed to ask the patient. HCPs recording ethnicity by observation method in one US report believed the data to be accurate as they felt they knew the local population (Regenstein and Sickler, 2006). Others may have simply seized the opportunity for a less confrontational route.

Reluctance to collect ethnicity data was encountered first hand during the course of the ethnicity data collection exercise I conducted at the CRN (reported in Chapter Nine). Research nurses felt uncomfortable asking the patients for ethnicity information believing it to be too sensitive and because they believed the patients were too fragile and already overwhelmed having received a cancer diagnosis coupled with treatment and trial information.

Fear of causing the patients offence or of being challenged by a patient was a worry for some staff as demonstrated in my survey of healthcare professional's perspectives of ethnicity data collection reported in Chapter Six. One healthcare professional felt it would be particularly offensive to ask younger ethnic minorities who were more likely to be British born for ethnicity information. A hospital in the USA, identified by the ethnicity data collection literature review (Chapter Four), addressed the problem of staff recording large numbers of 'Not known' codes through quality monitoring and provided extra training as needed (Regenstein and Sickler, 2006).

10.1.4 Methods currently used by researchers due to poor ethnicity data

The absence of complete ethnicity data has meant researchers have had to adopt other methodologies such as 1) using proxy indicators where available e.g. Country of Birth, 2) use of name recognition software such as Nam Pehchan and SANGRA where use is limited to South Asians, 3) data linkage, 4) sensitivity analyses and 5) use of multiple imputation or 6) conduct studies targeting specific populations (Wild et al, 2007; Cummins et al, 1999; Nanchahal et al, 2001; Ali et al, 2010; Fischbacher et al, 2007; Downing et al, 2011).

In a report of cancer incidence by ethnic group (2009) several permutations of sensitivity analyses were performed to assign ethnicity to the 24% of patients with missing ethnicity, but crude techniques like this can lead to results that are

problematic to interpret. Ryan et al (2012), reported inadequacies in name recognition software and the use of census data when performing analyses of cancer registry data. Downing et al (2011) opted for the more complex multiple imputation in their exploration of the relationship between ethnicity, breast cancer incidence and survival. This method was also used by the Office for National Statistics in a study of infant mortality by ethnic group. However, multiple imputation is based upon untestable assumptions, in cases where ethnicity is not missing at random (e.g. concentrated in particular ethnic groups) multiple imputation is deemed inappropriate (Marshall et al, 2010).

10.1.5 Training and highlighting the usefulness of ethnicity data

Use of the data to identify health needs and improve services is needed to showcase the utility of these data to patients and HCP alike as both parties felt collected data were not being used. The problem is that the data are not used in routine reports because it is so poor, and it is not collected because its value has not being sufficiently demonstrated. Staff training and promoting awareness of the importance of ethnicity data and its uses are essential in order to improve ethnicity data collection in healthcare. Training packages should promote the use of standardised ethnicity questions and provide standardised explanations to the questions frequently asked by patients, such as the purpose of collection. As reported in the literature review of ethnicity data collection in Chapter Four, both the USA and Scotland have toolkits which are freely available and provide

downloadable training materials (Health Research and Education Trust, 2007 and Health Scotland, 2005).

In the long run, using the ethnicity data we have in healthcare and government reports will highlight the flaws and may be the incentive required to improve ethnicity data collection.

10.2 Participation in RCTs

The systematic review of interventions to increase ethnic minority participation into RCTs highlighted the lack of research in this area from the UK (Chapter Seven). The USA literature revealed awarding grants to sites, the use of financial incentives and community representatives to be the most effective in recruiting ethnic minority participants. Financial incentives were also shown to aid retention, particularly where payments were staggered across the duration of the trial as demonstrated in the study reported by Germino et al (2011). The issues faced in the USA are very different to the UK with much of the literature focused upon addressing the lack of trust in researchers that African Americans feel following the infamous Tuskegee experiment.

There are important differences between the UK and USA healthcare systems. Care in the UK National Health Service is freely available, meaning that patients have access to treatment even when they are not in a clinical trial. Conversely, in the USA health insurance is usually provided by employers. Individuals out of work or those

in low paid jobs with either low or no medical cover may have very limited access to treatment.

The offer of free healthcare and/or the provision of financial incentives were found to be useful in increasing minority recruitment to clinical trials in the USA. In the UK, ethics committees only allow patients to receive payment for participating in research limited to costs incurred as a result of participation e.g. travel, food, loss of earnings, or childcare. Any additional payments are deemed to be 'inducement' and thought to lead to patients not considering the risks of the trial sufficiently or withholding information from health care professionals that would prevent them from taking part (Beckford and Broome, 2007). Australia has a similar policy, where payment above and beyond costs incurred or a fixed amount felt to be a reasonable estimate of out-of-pocket expenses is not allowed. The Australia Ethics National Statement states:

> *'Those who are economically disadvantaged might be exploited if payment were such that it provided what, in effect, would be a perverse incentive to take risks that they would otherwise not take, for example, payment that is greater than the current minimum wage'* (Australian Health Ethics Committee, 2009).

Interventions reported to be effective in improving the recruitment of ethnic minority participants identified by my review (Chapter Seven) included government grants to sites serving ethnic rich populations which allowed sites to invest in

additional resources. Community health workers and ethnicity matched researchers successfully overcame lack of trust issues in African American communities. Targeted mailshots to addresses in ethnic rich areas were reported to be the most cost effective method. Least successful strategies were found to be recruiting through health insurance companies, church events and health fairs.

10.3 Implications for the UK

Cancer clinical trials are now starting to routinely collect ethnicity data. Ethnicity questions were prospectively included in six trials currently being conducted at Warwick Clinical Trials Unit (three breast cancer, two head and neck, and one multiple myeloma trial). Initial results indicate that the majority of participants entering these trials are White. Ethnicity data from trials when teamed with Cancer Registry data and HES data will allow an assessment of the representation of ethnic minorities in clinical trials in the future.

Recruiting ethnic minorities into clinical trials also requires additional resources as demonstrated in the article by Cook et al (2010) where selected sites were awarded grants in order to facilitate the recruitment of African Americans. The funding was used in a variety of ways, the most successful methods included extra staffing permitting out of hours screening, transportation and parking payments, development of recruitment materials, refreshments, use of media advertising and mailshots targeting ethnic minority specific media channels, stores or areas of residence.

Additional barriers to participation, such as raising awareness of the need for research, the potential benefits of participating in a trial, the need for validated questionnaires in languages other than English and opportunities to request same sex healthcare professionals at assessments have been reported and may need to be considered when planning future interventions (Ford et al, 2005, Hussain-Gambles et al, 2004). Non-drug trials may also need to consider cultural issues and beliefs, such as modest dress and mixing of genders, i.e. in a study including physical exercise or group therapy.

However, the issue of participation and access to clinical trials is not restricted to ethnic minorities, other minority groups who don't speak English or don't understand, cannot read or write, who aren't educated and aren't aware of research and how research can be used to improve treatment/outcomes will experience similar inequities. Collecting data to monitor inequities in access to clinical trials for groups such as these is also a challenge. Indeed, the USA NIH revitalization Act 1993 is not restricted to ethnic minorities but instead mandates the inclusion of women and members of all minority groups and their subpopulations (NIH, 1993).

10.3.1 Examples from practice (insights from colleagues)

Experiences with inappropriate interpreters

Interpreters have been used in many hospitals within the NHS in the UK. However, difficulties experienced whilst using independent interpreters include age and

gender conflicts, such as a young Polish female patient with breast cancer allocated an elderly Ukrainian male interpreter.

However, there are issues surrounding the acceptability of relatives as interpreters. Clinicians are encouraged not to use relatives if avoidable, however, this can lead to problems when it is the patients wish to have a family member to interpret, and do not want the services/presence of a stranger. In South Asian communities, healthcare professional found most patients preferred to use relatives and were accompanied to appointments by an English speaking relative.

There are advantage of using interpreters over relatives because healthcare professionals have much more control of interpreters and can tell them exactly what to tell the patient, clinicians have much less power over relatives.

In one hospital, I learned they had stopped using interpreters due to cost issues and instead used a telephone service which cut costs but wasn't effective. Prior to this, the hospital would request an interpreter and he or she would be on standby until the patient was called. However, there are often long waits during busy clinics and the meter would be running for the entire duration. As it would be unethical to prioritise patients needing an interpreter to be seen first, interpreters were replaced by a dedicated telephone interpretation service. However, there were many problems with language mismatches, such as a Portuguese speaking patient from Brazil who was connected to a Portuguese speaking interpreter from Portugal which is different. The call was abandoned.

Time constraints

One clinician explained that it takes approximately 30 minutes to explain diagnosis and treatment information to a new cancer patient when the patient speaks English but has no medical knowledge. This time increased when discussing a trial as it may involve explaining what a trial is and why it needs to be conducted. A clinician may need to explain concepts such as equipoise, randomisation and possible treatment options as well as what treatments are available if they choose not to participate in the trial. This process is much more time-consuming with non-English speaking patients where awareness of research and trials is much lower.

Consent issues

A lead research nurse from Dorset described the complexities of taking consent from potential cancer trial participants and the need for the person taking consent to be reassured that the person giving the consent had understood the information. She described how particularly difficult this was when considering non-English speaking ethnic minorities. At times she wasn't sure if they could read the information leaflet and therefore didn't feel they were giving informed consent or that the consent process was what it should be. On several occasions she was asked by consultants to consent patients and refused because she didn't feel sure that consent could be given in a really informed way.

One consultant felt the most daunting thing about recruiting ethnic minorities was the clinician feeling that he or she cannot explain the study to the patient correctly,

and the difficulties of doing this through an interpreter. Most patient information leaflets are not available in languages other than English and there are so few ethnic minority research nurses. The consultant went onto to say "I think it's effort, it's effort, it's effort, that's why" stating that he/she believed ethnic minorities to be under-represented in clinical trials because they were much harder to recruit at every step.

Use of translated patient information sheets and consent forms

A chief investigator of a national phase III cancer trial reported translating patient information sheets and consent forms into the two most common non-English ethnic minority languages within the catchment areas for the centres recruiting into the study (Urdu and Gujarati). These were sent to all participating centres in the UK, irrespective of ethnic minority population composition. However, the use of the translated patient information sheets was very low. The research nurses and doctors were not trained to be culturally sensitive and did not try to identify suitable patients or use the patient information sheets. An estimated eight eligible patients who were literate in Urdu or Gujarati were identified from a total of 68 UK centres and only two of these patients consented to participate in the trial. The initial cost of translating the information was large and increased each time an amendment was made, but the resulting recruitment was small. More recently, the same Chief Investigator has adopted a 'translate on demand' approach, providing patient information sheets and consent forms in any language required upon request. However, the 'translate on demand' approach has not proved to be

affective in practice so far (i.e. no requests) possibly due to the 'bother' factor, literacy levels in older ethnic minority groups and a lack of awareness of the initiative.

There is some evidence to suggest acculturation of cancer rates is occurring. This is the process where the risk of ethnic minorities developing disease changes to become similar to that of the host population. It has already been reported to be happening in breast cancer rates in the UK South Asian population (Farooq and Coleman, 2005). Migration studies have also shown this. A study of South Asians from India, Singapore, UK and the USA found the lowest rates of cancer in India and higher rates in Indian immigrants in Singapore, UK and the USA (Rastogi et al, 2008).

10.4 Recommendations

To address the poor quality of ethnicity data collection, the collection of self-reported ethnicity data should be mandated the first time a patient comes into contact with the NHS, preferably in primary care and such data should then be linked through to secondary care and verified at subsequent appointments where needed. Providing healthcare professionals with training and access to ethnicity monitoring toolkits with standardised questions and rationale would help combat barriers to ethnicity data collection at the healthcare professional level. Demonstrating the value of the collected data in reports and using it to improve

services would also encourage ethnicity data collection and provision and would highlight the deficiencies in the data.

Recruitment of ethnic minority groups in clinical trials could be improved by mandating the inclusion of a representative sample of ethnic minorities (representative of the population and/or incidence of the condition of interest) in clinical trials, similar to the model demonstrated by the USA's NIH Revitalization Act 1993 and the successful NCRN who were established with the aim of doubling the recruiting of all cancer patients into clinical trials from 3.75% to 7.5% of incident cases (NCRN, 2006). Research in the USA has shown the effectiveness of patient navigators, cultural brokers or link workers to facilitate the recruitment of ethnic minority patients into clinical trials. The role would involve using a member of the same community as the target group of patients to minimise cultural and language barriers. The facilitators should have knowledge of clinical trials in general and of the clinical trials available and be able to give patients the fullest information possible enabling them to make an informed decision.

Patients who are not proficient in English were encouraged by research staff at the CRN to attend appointments with an English speaking relative or friend (Chapter Nine). Continuation of this recommendation would overcome the difficulty of some patients not even being approached to enter an RCT and help minimise problems related to matching languages and dialects of patients to interpreters. However, independent interpreters may need to be available as it is not certain that a relative could provide all the necessary information for 'fully informed consent'.

10.5 Conclusions

A change in local and national policy to mandate the collection of ethnicity data in primary care and linked through to other healthcare service providers is needed.

Healthcare professionals should be familiarised with the Equality Act 2010 in training exercises or at induction to raise awareness of the need for ethnicity data collection and how this data will be used to improve services. Unused data were reported to be a disincentive to both healthcare professionals and patients alike (Iqbal et al, 2012, Fulton, 2010). Training should also emphasise the importance of self-reported ethnicity as well as providing standardised: 1) rationale, 2) wording of questions, 3) response categories offered and 4) answers and explanations to frequently asked questions. Only when we have better ethnicity data will we be able to assess the true extent of inequities in access to healthcare and clinical trials. Investigations of barriers to ethnic minority participation in clinical trials are needed before we identify appropriate strategies to combat them.

Reducing health inequalities, improving access to clinical trials and tailoring current services to meet the needs of the UK's ethnic minority population is dependent on having accurate ethnicity data. Patients need to feel confident that their data will be handled confidentially and used to improve services (Johnson 2012, Fulton 2010). Without improvement in the recording and use of this information we will continue to remain blind to the size of the problem. Collecting ethnicity data is of no use if the data are not used to target resources and reduce inequalities (Raleigh,

2008). In the words of Fulton (2010) *'health equality is not possible without ethnic monitoring'*.

There has been a noticeable change in ethnicity data being collected since the start of this thesis in 2006, the CanEth report in 2009 and the papers I have published. Through data linkage with HES and ethnicity data collection on new registrations, ethnicity data of cancer patients are now becoming available through the National Cancer Data Repository (NCIN et al, 2010). Researchers using the data will identify the flaws in the ethnicity data collected and drive improvement and change.

The outcomes from each of the research components in this thesis highlight challenges and solutions to the collection and use of ethnicity data in the UK, completing the research cycle from initial empirical observation to implementation in practice.

References

1948. British Nationality Act. United Kingdom.

1998. 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 [Online]. AIR Alert. Available: http://www.fsu.edu/-air/alert6.htm.

2000. Race Relations (Ammendment) Act 2000. UK.

2001. NIH policy on reporting race and ethnicity data: subjects in clinical research [Online]. Available: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-01-053.html.

Adebayo, C & Mitchell, P 2004. Patient profiling. Liverpool: Central Liverpool Primary Care Trust.

Alcalai, R, Ben-Yehuda, D, Ronen, I & Paltiel, O 2003. Ethnicity and prognosis in acute myeloid leukemia. American Journal of Hematology, 72, 127-134.

Ali, R, Barnes, I, Kan, SW & Beral, V 2010. Cancer incidence in British Indians and British whites in Leicester, 2001-2006. British Journal of Cancer, 103, 143-148.

Aspinall, P 1998. Describing the "white" ethnic group and its composition in medical research. Soc Sci Med, 47, 1797-808.

Aspinall, PJ 2000. The mandatory collection of data on ethnic group of inpatients: experience of NHS trusts in England in the first reporting years. Public Health, 114, 254-9.

Aspinall, PJ 2010. Concepts, terminology and classifications for the "mixed" ethnic or racial group in the United Kingdom. J Epidemiol Community Health, 64, 557-60.

Aspinall, PJ & Jacobson, B 2004. Ethnic disparities in health and health care: A focused review of the evidence and selected examples of good practice.

Austin, LT, Ahmad, F, Mcnally, MJ & Stewart, DE 2002. Breast and cervical cancer screening in Hispanic women: a literature review using the health belief model. Womens Health Issues, 12, 122-8.

Australian Health Ethics Committee 2009. Using the national statement 1: Payments to participants in research, particularly clinical trials. Australia.

Baider, L & Sarell, M 1983. Perceptions and causal attributions of Israeli women with breast cancer concerning their illness - The effects of ethnicity of health and religiosity. PSYCHOTHERAPY AND PSYCHOSOMATICS, 39, 136-143.

Baker, DW, Cameron, KA, Feinglass, J, Georgas, P, Foster, S, Pierce, D, Thompson, JA & Hasnain-Wynia, R 2005. Patients' attitudes toward health care providers collecting information about their race and ethnicity. Journal of General Internal Medicine, 20, 895-900.

Baker, DW, Hasnain-Wynia, R, Kandula, NR, Thompson, JA & Brown, ER 2007. Attitudes toward health care providers, collecting information about patients' race, ethnicity, and language. Medical Care, 45, 1034-1042.

Beckford, L & Broome, MR 2007. Ethics and the payment of research subjects. Psychiatry, 6, 83-85.

Begum, G, Bryan, R.T., Dunn, J.A., Bathers, S., Wallace, D.M.A. The effect of material deprivation on survival in bladder cancer. European Association of Urology, 2002 Birmingham. European Urology.

Begum, G, Dunn, JA, Bryan, RT, Bathers, S, Wallace, DM & West Midlands Urological Research, G 2004. Socio-economic deprivation and survival in bladder cancer. BJU Int, 94, 539-43.

Bhopal, R 2004. Glossary of terms relating to ethnicity and race: for reflection and debate. J Epidemiol Community Health, 58, 441-5.

Bhopal, R 2007. Ethnicity, Race, and Health in Multicultural Societies: Foundations for Better Epidemiology, Public Health, and Health Care, Oxford University Press.

Bowen, R, Duffy, S, Ryan, D, Hart, I & Jones, J 2008. Early onset of breast cancer in a group of British black women. BRITISH JOURNAL OF CANCER, 98, 277-281.

Bradford, B 2006. Who are the 'Mixed' ethnic group? London.

Braunholtz, DA, Edwards, SJ & Lilford, RJ 2001. Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect". J Clin Epidemiol, 54, 217-24.

Brawley, OW 1998. The study of untreated syphilis in the negro male. Int J Radiat Oncol Biol Phys, 40, 5-8.

Breast Cancer Care Policy Briefing 2005. Same difference: breast awareness is for everyone.

Britton, A, Mckee, M, Black, N, Mcpherson, K, Sanderson, C & Bain, C 1999. Threats to applicability of randomised trials: exclusions and selective participation. J Health Serv Res Policy, 4, 112-21.

Brown, SD, Lee, K, Schoffman, DE, King, AC, Crawley, LM & Kiernan, M 2012. Minority recruitment into clinical trials: Experimental findings and practical implications. Contemporary Clinical Trials, 33, 620-623.

Carr-Hill, R, Passingham, S, Wolf, A & Kent, N 1996. Lost Opportunities: The Language Skills of Linguistic Minorities in England and Wales. London.

Centers for Disease Control and Prevention 1999. Reporting race and ethnicity data--National Electronic Telecommunications System for Surveillance, 1994-1997. MMWR - Morbidity & Mortality Weekly Report, 48, 305-12.

NHS Health and Social Care Information Centre, 2006. Health Survey for England 2004: The health of minority ethnic groups.

Chattar-Cora, D, Onime, GD, Valentine, IS, Cudjoe, E & Rivera, L 2000. Colorectal cancer in a multi-ethnic urban group: its anatomical and age profile. International Surgery, 85, 137-42.

Commission for Racial Equality 2002. Ethnic monitoring: A guide for public authorities.

Commission for Racial Equality 2007. Ethnic minorites in Great Britain. Factfile 2. In: COMMISSION FOR RACIAL EQUALITY (ed.). London.

Cook, ED, Arnold, KB, Hermos, JA, Mccaskill-Stevens, W, Moody-Thomas, S, Probstfield, JL, Hamilton, SJ, Campbell, RD, Anderson, KB & Minasian, LM 2010. Impact of supplemental site grants to increase African American accrual for the Selenium and Vitamin E Cancer Prevention Trial. Clinical Trials, 7, 90-9.

Cook, ED, Moody-Thomas, S, Anderson, KB, Campbell, R, Hamilton, SJ, Harrington, JM, Lippman, SM, Minasian, LM, Paskett, ED, Craine, S, Arnold, KB & Probstfield, JL 2005. Minority recruitment to the Selenium and Vitamin E Cancer Prevention Trial (SELECT). Clinical Trials, 2, 436-42.

Cooper, H, Lester, H, Wilson, S, Freemantle, N & Riley, H 2003. Representation of South Asian people in randomised trials: ethnic origin need not be a barrier to participation. BMJ, 327, 394-5.

Cummins, C, Winter, H, Cheng, KK, Maric, R, Silcocks, P & Varghese, C 1999. An assessment of the Nam Pehchan computer program for the identification of names of south Asian ethnic origin. Journal of Public Health Medicine, 21, 401-6.

Dein, S 2004. Explanatory models of and attitudes towards cancer in different cultures. Lancet Oncol, 5, 119-24.

Department of Health 1999. Health Survey for England 1999: The health of minority ethnic groups.

Department of Health 2001. Collecting ethnic category data: Guidance and training material for implementation of the new ethnic categories.

Department of Health 2005. A practical guide to ethnic monitoring in the NHS and social care.

Douglas, A, Bhopal, RS, Bhopal, R, Forbes, JF, Gill, JMR, Lawton, J, Mcknight, J, Murray, G, Sattar, N, Sharma, A, Tuomilehto, J, Wallia, S, Wild, SH & Sheikh, A 2011. Recruiting South Asians to a lifestyle intervention trial: experiences and lessons from PODOSA (Prevention of Diabetes & Obesity in South Asians). Trials [Electronic Resource], 12, 220.

Downing, A, West, RM, Gilthorpe, MS, Lawrence, G & Forman, D 2011. Using routinely collected health data to investigate the association between ethnicity and breast cancer incidence and survival: what is the impact of missing data and multiple ethnicities? Ethnicity & Health, 16, 201-212.

Du, W, Mood, D, Gadgeel, S & Simon, MS 2009. An educational video to increase clinical trials enrollment among breast cancer patients. Breast Cancer Research & Treatment, 117, 339-47.

Ellis, A 2009. UK resident population by country of birth. London: Office for National Statistics.

Ethnic Minority Cancer Awareness Week 2008. Ethnic Minority Cancer Awareness Week Report 2008.

Farooq, S & Coleman, MP 2005. Breast cancer survival in South Asian women in England and Wales. Journal of Epidemiology and Community Health, 59, 402-406.

Fischbacher, C, Bhopal, R, Povey, C, Steiner, M, Chalmers, J, Mueller, G, Jamieson, J & Knowles, D 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.

Ford, JG, Howerton, MW, Lai, GY, Gary, TL, Bolen, S, Gibbons, MC, Tilburt, J, Baffi, C, Tanpitukpongse, TP, Wilson, RF, Powe, NR & Bass, EB 2008. Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. Cancer, 112, 228-42.

Ford, ME & Kelly, PA 2005. Conceptualizing and categorizing race and ethnicity in health services research. Health Services Research, 40, 1658-75.

Fulton, R 2010. Ethnic monitoring: is health equality possible without it? A Race Equality Foundation Briefing Paper. London.

Gardener, D & Connolly, H 2005. Who are the 'Other' ethnic groups? London.

Gardi, M 2003. Ethnicity data protocols training presentation. Ministry of Health Manatu Hauora.

Gatineau, M & Mathrani, S 2011. Ethnicity and obesity in the UK. Perspectives in Public Health, 131, 159-60.

Gatineau, M & Mathrani, S 2011. Obesity and ethnicity. Oxford.

Germino, BB, Mishel, MH, Alexander, GR, Jenerette, C, Blyler, D, Baker, C, Vines, AI, Green, M & Long, DG 2011. Engaging African American breast cancer survivors in an intervention trial: culture, responsiveness and community. Journal of Cancer Survivorship, 5, 82-91.

Gill Ps, Kai J, Bhopal Rs & Wild S 2007. Black and minority ethnic groups. Healthcare Needs Assessment.

Godden, S, Ambler, G & Pollock, AM 2010. Recruitment of minority ethnic groups into clinical cancer research trials to assess adherence to the principles of the Department of Health Research Governance Framework: national sources of data and general issues arising from a study in one hospital trust in England. J Med Ethics, 36, 358-62.

Gordon-Dseagu, V 2008. Cancer and health inequalities: An introduction to current evidence.

Gotay, CC & Holup, J 2004. Ethnic identities and lifestyles in a multi-ethnic cancer patient population. Pacific Health Dialog, 11, 191-8.

Green, G, Bradby, H, Chan, A & Lee, M 2006. "We are not completely Westernised": dual medical systems and pathways to health care among Chinese migrant women in England. Soc Sci Med, 62, 1498-509.

Green, J & Thorogood, N 2004. Qualitative Methods for Health Research, Sage publications.

Harris, KJ, Ahluwalia, JS, Catley, D, Okuyemi, KS, Mayo, MS & Resnicow, K 2003. Successful recruitment of minorities into clinical trials: The Kick It at Swope project. Nicotine & Tobacco Research, 5, 575-84.

Hasnain-Wynia, R & 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, 1501-1518.

Hasnain-Wynia, R, Pierce, D & Pittman, M 2004. Who, when, and how: The current state of race, ethnicity, and primary language data collection in hospitals. Health Research and Educational Trust and The Commonwealth Fund.

Hasnain-Wynia, R, Taylor-Clark, K & Anise, A 2011. Collecting Race, Ethnicity, and Language Data to Identify and Reduce Health Disparities: Perceptions of Health Plan Enrollees. Medical Care Research and Review, 68, 367-381.

Health, DO 2011. Improving Outcomes: A Strategy for Cancer. In: HEALTH, D. O. (ed.).

Health Research and Education Trust. 2007. HRET Disparities Toolkit: A toolkit for collecting race, ethnicity and primary language information for patients [Online].

Health Scotland. 2005. Ethnic Monitoring Tool [Online]. Available: http://www.isdscotland.org/isd/files/ETHNIC%20MONITORING%20TOOL.pdf.

Helman, C 2000. Culture, Health and Illness, Hodder Arnold.

Hes Online. 2004. How good is HES ethnicity coding and where do the problems lie? [Online]. Available: http://www.hesonline.nhs.uk.

Hes Online 2011. How good is HES ethnic coding and where do the problems lie? In: CENTRE, T. H. A. S. C. I. (ed.).

Hiller, L, Radley, S, Mann, CH, Radley, SC, Begum, G, Pretlove, SJ & Salaman, JH 2002. Development and validation of a questionnaire for the assessment of bowel and lower urinary tract symptoms in women. BJOG, 109, 413-23.

Home Office 2010. Equality Act 2010. In: OFFICE, H. (ed.).

Hord, M, Smith, T, Culbert, S, Frankel, L & Pinkel, D 1996. Ethnicity and cure rates of Texas children with acute lymphoid leukemia. CANCER, 77, 563-569.

Horowitz, CR, Brenner, BL, Lachapelle, S, Amara, DA & Arniella, G 2009. Effective recruitment of minority populations through community-led strategies. American Journal of Preventive Medicine, 37, S195-200.

Hussain-Gambles, M, Leese, B, Atkin, K, Brown, J, Mason, S & Tovey, P 2004. Involving South Asian patients in clinical trials. Health Technology Assessment, 8 (42).

Information Services Division Scotland 2010. Patient profiling pilot project. Improving ethnic data collection for equality and diversity monitoring. Scotland.

Jack, RH, Linklater, KM, Hofman, D, Fitzpatrick, J & Moller, H 2006. Ethnicity coding in a regional cancer registry and in Hospital Episode Statistics. BMC Public Health, 6:8.

Johnson, M 2012. Problems and barriers in the collection of ethnicity and migration related data. In: INGLEBY, D., KRASNIK, A., LORANT, V. & RAZUM, O. (eds.) Health inequalities and risk factors among migrants and ethnic minorities. Garant: Antwerp/Apeldoorn.

Johnson, MRD 2000. Black and minority ethnic groups in England: the second health and lifestyles survey. London.

Johnson, MRD, Hardman, C & Jones, A 1993. Developing the implementation of ethnic monitoring: Final report to the Home Office In: EMDP/CRER (ed.). Warwick University.

Karanth, M, Begum, G, Cook, M, Lawson, S, Porter, C, Lister, N, Gray, J, Darbyshire, P, Craddock, C & Mahendra, P 2006. Increased acute GvHD and higher transplant-related mortality in non-caucasians undergoing standard sibling allogeneic stem cell transplantation. Bone Marrow Transplant, 37, 419-23.

Karjalainen, S & Palva, I 1989. Do treatment protocols improve end results? A study of survival of patients with multiple myeloma in Finland. BMJ, 299, 1069-72.

Kelly, JJ, Chu, SY, Diaz, T, Leary, LS & Buehler, JW 1996. Race/ethnicity misclassification of persons reported with AIDS. The AIDS Mortality Project Group and The Supplement to HIV/AIDS Surveillance Project Group. Ethnicity & Health, 1, 87-94.

Kiernan, M, Phillips, K, Fair, JM & King, AC 2000. Using direct mail to recruit Hispanic adults into a dietary intervention: an experimental study. Annals of Behavioral Medicine, 22, 89-93.

Killien, M, Bigby, JA, Champion, V, Fernandez-Repollet, E, Jackson, RD, Kagawa-Singer, M, Kidd, K, Naughton, MJ & Prout, M 2000. Involving minority and underrepresented women in clinical trials: the National Centers of Excellence in Women's Health. J Womens Health Gend Based Med, 9, 1061-70.

King's Fund 2006. Access to health care and minority ethnic groups-Briefing.

Kumarapeli, P, Stepaniuk, R, De Lusignan, S, Williams, R & Rowlands, G 2006. Ethnicity recording in general practice computer systems. J Public Health (Oxf), 28, 283-7.

Lawrence, G, Kearins, O, Lagord, C, Cheung, S, Sidhu, J & Sagar, C 2011. The second All Breast Cancer Report.

London Health Observatory 2003. Diversity counts. Ethnic health intelligence in London: the story so far.

Ma, IWY, Khan, NA, Kang, A, Zalunardo, N & Palepu, A 2007. Systematic review identified suboptimal reporting and use of race/ethnicity in general medical journals. Journal of Clinical Epidemiology, 60, 572-578.

Macmillan. 2011. Who gets cancer [Online]. Macmillan Cancer Information. Available: (http://www.macmillan.org.uk/Cancerinformation/Aboutcancer/Whogetscancer.aspx [Accessed 06/Mar/2012 2012].

Macpherson, W 1999. The Stephen Lawrence Inquiry.

Marshall, A, Altman, DG, Royston, P & Holder, RL 2010. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. BMC Med Res Methodol, 10, 7.

Martin, MA, Swider, SM, Olinger, T, Avery, E, Lynas, CMT, Carlson, K & Rothschild, SK 2011. Recruitment of Mexican American adults for an intensive diabetes intervention trial. Ethnicity & Disease, 21, 7-12.

Mason, S, Hussain-Gambles, M, Leese, B, Atkin, K & Brown, J 2003. Representation of South Asian people in randomised clinical trials: analysis of trials' data. BMJ, 326, 1244-5.

Matthews, HW 1995. Racial, ethnic and gender differences in response to medicines. Drug Metabol Drug Interact, 12, 77-91.

Mhurchu, CN, Blakely, T, Funaki-Tahifote, M, Mckerchar, C, Wilton, J, Chua, S & Jiang, Y 2009. Inclusion of indigenous and ethnic minority populations in intervention trials: challenges and strategies in a New Zealand supermarket study. Journal of Epidemiology & Community Health, 63, 850-5.

Murray, S & Buller, AM 2007. Exclusion on grounds of language ability--a reporting gap in health services research? J Health Serv Res Policy, 12, 205-8.

Nanchahal, K, Mangtani, P, Alston, M & Silva, ID 2001. 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, 23, 278-285.

National Cancer Intelligence Network 2009. Cancer incidence and survival by major ethnic group, England, 2002-2006. London.

National Statistics 2003. Ethnic group statistics: A guide for the collection and classification of ethnicity data.

Ncin, U, Ons,. 2010. National Cancer Research Network,. Available: http://www.ncin.org.uk/collecting_and_using_data/national_cancer_data_repository/ [Accessed 20/Aug/2013 2013].

Network, NCI 2009. Cancer incidence and survival by major ethnic group, England, 2002-2006.

NHS Care Records Service. 2011. Summary Care Records [Online]. Department of Health. Available: http://www.nhscarerecords.nhs.uk/index.html [Accessed 19/08/2011 2011].

Observatory, TM. 2013. Migration to the UK: Asylum [Online]. University of Oxfod. Available: http://www.migrationobservatory.ox.ac.uk/briefings/migration-uk-asylum [Accessed 20-Aug-2013].

O'Donnell, PH & Dolan, ME 2009. Cancer pharmacoethnicity: ethnic differences in susceptibility to the effects of chemotherapy. Clin Cancer Res, 15, 4806-14.

Office for National Statistics. 2005. Focus on Ethnicity and Identity: Religion: 7 in 10 identify as Christian [Online]. London: Office for National Statistics. Available: www.statistics.gov.uk/cci/nugget.asp?id=460 [Accessed 16 October 2009].

Office for National Statistics. 2006. Focus on ethnicity and identity: Age/Sex Distribution: Non-White groups are younger [Online]. Available: www.statistics.gov.uk/cci/nugget.asp?id=456 [Accessed 16 October 2009].

Office for National Statistics. 2011. Ethnicity and National Identity in England and Wales 2011 [Online]. Available: http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-authorities-in-england-and-wales/rpt-ethnicity.html#tab-Ethnicity-in-England-and-Wales [Accessed 20/Aug/2013].

Office for National Statistics 2013. Harmonised Concepts and Questions for Social Data Sources: Primary Standards-Ethnic Group. 3.2 ed.

Owen, D 1996. Size, structure and growth of the ethnic minority populations. In: COLEMAN, D. & SALT, J. (eds.) Ethnicity in the 1991 Census Volume One. Demographic characteristics of the ethnic minority populations. London: HMSO.

Perkins, P, Cooksley, C & Cox, J 1996. Breast cancer - Is ethnicity an independent prognostic factor for survival? CANCER, 78, 1241-1247.

Pinnock, K. 2009. Mixed race 'fastest growing minority'. 12/Aug/2009.

Pringle, M & Rothera, I 1996. Practicality of recording patient ethnicity in general practice: Descriptive intervention study and attitude survey. British Medical Journal, 312, 1080-1082.

Prostate Cancer UK. 2013. African and African Caribbean men [Online]. Available: http://prostatecanceruk.org/we-can-help/african-caribbean-communities [Accessed 08/Aug/2012 2013].

Race for Health 2006. Lambeth PCT peer review.

Race for Health 2007. Ethnic Monitoring in General Practice – communicating the benefits.

Raleigh, V 2008. Collection of data on ethnic origin in England. BRITISH MEDICAL JOURNAL, 337:a1107, doi: 10.1136/bmj.a1107.

Rastogi, T, Devesa, S, Mangtani, P, Mathew, A, Cooper, N, Kao, R & Sinha, R 2008. Cancer incidence rates among South Asians in four geographic regions: India, Singapore, UK and US. Int J Epidemiol, 37, 147-60.

Regenstein, M & Sickler, D 2006. Race, ethnicity, and language of patients. Hospital practices regarding collection of information to address disparities in health care. The Robert Wood Johnson Foundation and National Public Health and Hospital Institute.

Robb, KA, Solarin, I, Power, E, Atkin, W & Wardle, J 2008. Attitudes to colorectal cancer screening among ethnic minority groups in the UK. BMC Public Health, 8, 34.

Ross, S, Grant, A, Counsell, C, Gillespie, W, Russell, I & Prescott, R 1999. Barriers to participation in randomised controlled trials: A systematic review. JOURNAL OF CLINICAL EPIDEMIOLOGY, 52, 1143-1156.

Rothwell, PM 2005. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet, 365, 82-93.

Ryan, R, Vernon, S, Lawrence, G & Wilson, S 2012. Use of name recognition software, census data and multiple imputation to predict missing data on ethnicity: application to cancer registry records. BMC Medical Informatics and Decision Making, 12:3.

Sangowawa, O & 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, 106-108.

Schneider, H & Palmer, N 2002. Getting to the truth? Researching user views of primary health care. Health Policy Plan, 17, 32-41.

Sheth, T, Nargundkar, M, Chagani, K, Anand, S, Nair, C & Yusuf, S 1997. Classifying ethnicity utilizing the Canadian Mortality Data Base. Ethnicity & Health, 2, 287-95.

Sproston, K & Mindell, J 2006. Health survey of England 2004. Volume 1: The health of minority ethnic groups.

Swallen, KC, West, DW, Stewart, SL, Glaser, SL & Horn-Ross, PL 1997. Predictors of misclassification of Hispanic ethnicity in a population-based cancer registry. Annals of Epidemiology, 7, 200-6.

Szczepura, A 2005. Access to health care for ethnic minority populations. Postgrad Med J, 81, 141-7.

Szczepura, A, Price, C & Gumber, A 2008. Breast and bowel cancer screening uptake patterns over 15 years for UK south Asian ethnic minority populations, corrected for differences in socio-demographic characteristics. BMC Public Health, 8, 346.

United Nations Statistics Division 2003. Ethnicity: A Review of Data Collection and Dissemination. Social and Housing Statistics Section, Demographic and Social Statistics Branch.

Us Census Bureau. Available: http://www.census.gov/2010census/data/ [Accessed 05/Jul/2013.

Wallace, DMA, Bryan, RT, Dunn, JA, Begum, G & Bathers, S 2002. Delay and survival in bladder cancer. Bju International, 89, 868-878.

Ward, LC, Slegg, GP, Dunn, JA, Harvey, P, Hallissey, MT, Hyde, C, Begum, G, Harris, S & Kerr, DJ 2000. A systematic review reporting doctors' and patients' attitudes toward participation in clinical research trials. NHS National Cancer R&D Programme Project.

Weinick, RM, Caglia, JM, Friedman, E & Flaherty, K 2007. Measuring racial and ethnic health care disparities in Massachusetts. Health Affairs, 26, 1293-1302.

Wendler, D, Kington, R, Madans, J, Van Wye, G, Christ-Schmidt, H, Pratt, LA, Brawley, OW, Gross, CP & Emanuel, E 2006. Are racial and ethnic minorities less willing to participate in health research? PLoS Medicine, 3(2)(pp 0201-0210).

White, A 2002. Social Focus in brief: Ethnicity. Office for National Statistics. London.

Wild, SH, Fischbacher, C, Brock, A, Griffiths, C & Bhopal, R 2007. Mortality from all causes and circulatory disease by country of birth in England and Wales 2001-2003. Journal of Public Health, 29, 191-198.

Williams, S 1999. Dental services for the Bangladeshi community. British Dental Journal, 186, 517-521.

Yasuda, SU, Zhang, L & Huang, SM 2008. The role of ethnicity in variability in response to drugs: focus on clinical pharmacology studies. Clin Pharmacol Ther, 84, 417-23.

List of Appendices

Appendix 1: Ethics approval for the CanEth project

30/07

07/Q2707/33

NHS
West Midlands
South Birmingham Research Ethics Committee
27 Highfield Road
Edgbaston
Birmingham B15 3DP
Rosa.Downing@westmidlands.nhs.uk
Telephone: 0121 245 2534 & 2538
Facsimile: 0121 245 2535

Chairman: Mr R K Vohra
Administrator: Mrs R M Downing

Ref: RMD
Date: 26 March 2007

Professor Janet A Dunn
Professor of Clinical Trials & Head of Cancer Trials
University of Warwick
The Medical School
Gibbet Hill Road
Coventry
CV4 7AL

Dear Professor Dunn

Full title of study: **Improving ethnic data collection for statistics of cancer incidence, management, mortality and survival in the UK**
REC reference number: **07/Q2707/33**

Thank you for your letter dated 19 March 2007, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information was considered by the Vice Chairman.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Ethical review of research sites

The Committee has not yet been notified of the outcome of any site-specific assessment (SSA) for the research site(s) taking part in this study. The favourable opinion does not therefore apply to any site at present. We will write to you again as soon as one Research Ethics Committee has notified the outcome of a SSA. In the meantime no study procedures should be initiated at sites requiring SSA.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
Application	5.3	
Investigator CV		19 October 2006
Protocol	2.0	22 January 2007
Covering Letter		19 March 2007
Letter from Sponsor		20 October 2006
Interview Schedules/Topic Guides	1.0	19 October 2006
Questionnaire: Health Care Professionals	4.0	17 January 2007
Letter of invitation to participant	2.1	19 March 2007
GP/Consultant Information Sheets	2.0	22 January 2007
Participant Information Sheet: Community	2.1	19 March 2007
Participant Consent Form: For GP participants	2.0	22 January 2007
Participant Consent Form: Community	2.1	19 March 2007
Evidence of Insurance		02 August 2006
Evidence of Insurance		03 August 2006
Invitation to participate in Survey	4.0	17 January 2007
Survey of ethnicity data collection for cancer statistics	4.2	19 March 2007
Letter from Mr P Haezewindt, Cancer Research UK regarding Funding		10 July 2006
Grant Award Letter from Dr C Moore		18 August 2006
Response to Mr P Haezewindt		14 July 2006

R&D approval

The study should not commence at any NHS site until the local Principal Investigator has obtained final approval from the R&D office for the relevant NHS care organisation.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

07/Q2707/33 **Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

Mr A D Mayer
Vice Chair

Enclosures: Standard approval conditions SL-AC2
Site approval form

Copy to: University of Warwick

Appendix 2: Ethnicity, Health and Diversity questionnaire

Ethnicity, Health and Diversity

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				Very important
1	2	3	4	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	No / Yes / Not known	Country of birth	No / Yes / Not known
Race	No / Yes / Not known	Country of origin	No / Yes / Not known
Religion	No / Yes / Not known	Language	No / Yes / Not known
Patient name (i.e. for use with name recognition software)	No / Yes / Not known		
Other	No / Yes / Not known		

If other please give details

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. Please return it to:
Gulnaz Begum, Rm. B-038, Medical School Building, University of Warwick, Coventry, CV4 7AL
Would you like to receive a copy of the final report: Yes / No

Appendix 3: Reason for Refusal and Ethnicity Form

Reason for Refusal and Ethnicity Form

I agree to complete the form ☐

I do not agree to complete the form (please return form in the pre-paid envelope) ☐

Date completed ☐☐ ☐☐☐ ☐☐☐☐
d d m o n y e a r

Age group (years): Under 40 ☐ 40-50 ☐ 50-60 ☐ 60-70 ☐ Over 70 ☐

Gender: Male ☐ Female ☐

1. Reason for declining entry to a trial

I decided against it because

2. Type of cancer (for example breast, colorectal, lung, skin)

3. Trial name

4. What is your ethnic group?
(Choose one section from A to E, then tick the appropriate box. If 'Other background' is selected please provide further details in the 'please specify' text box)

A. WHITE
- British ☐
- Irish ☐
- Other White background ☐ please specify

B. MIXED
- White and Black Caribbean ☐
- White and Black African ☐
- White and Asian ☐
- Other Mixed background ☐ please specify

C. ASIAN or ASIAN BRITISH
- Indian ☐
- Pakistani ☐
- Bangladeshi ☐
- Other Asian background ☐ please specify

D. BLACK or BLACK BRITISH
- Caribbean ☐
- African ☐
- Other Black background ☐ please specify

E. CHINESE or OTHER ETHNIC GROUP
- Chinese ☐
- Other ethnic group ☐ please specify

5. What is your religion? (Tick only one box)

- ☐ None
- ☐ Christian (including Church of England, Catholic, Protestant and all other Christian denominations)
- ☐ Buddhist
- ☐ Hindu
- ☐ Jewish
- ☐ Muslim
- ☐ Sikh
- ☐ Any other religion, please specify

6. What is your country of birth?

- ☐ England
- ☐ Wales
- ☐ Scotland
- ☐ Northern Ireland
- ☐ Republic of Ireland
- ☐ Elsewhere, if yes then provide country and years living in UK — Country — Years

7a. What is your first language?

7b. If given the choice what language would you feel most comfortable speaking with your doctor or nurse in?

8. How would you rate your ability to speak and understand English?

☐ Excellent ☐ Good ☐ Fair ☐ Poor ☐ Not at all

Thank you for your help

Please post back your form in the pre-paid envelope provided

Appendix 4: Conference posters

Ethnic Minorities in Oncology Clinical Trials

Gulnaz Begum and Janet Dunn
Warwick Clinical Trials Unit, The University of Warwick, UK

THE UNIVERSITY OF WARWICK

Background

- Ethnic minorities are reported to be under-represented in cancer clinical trials[1,2]
- The USA has sought to address this problem by mandating that a set percentage of patients recruited into NIH trials should be from minority groups[3]; however ethnic groups differ greatly between the USA and UK, with more diversity within the UK
- Barriers to recruitment include language within the informed consent process, lack of awareness of clinical trials, geographical location of participating centres and cultural changes in health beliefs[4,5]
- 270,000 new cases of cancer were diagnosed in the UK in 2001; 15% were breast cancer and 13% were bowel or colorectal cancers[6]
- 4.5 million people (8% of the population) in 2001 within the UK were defined as being from an ethnic minority group; in certain areas in the UK this number is as high as 30-40%. South Asians are the largest ethnic minority group, 2.7% of the total population (table 1)[7,8]

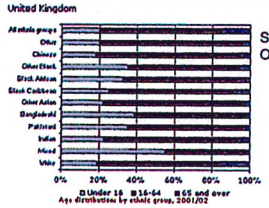
Table 1: The UK population: by ethnic group, April 2001

	Total population		Minority Ethnic population
Ethnic group	n	%	%
White	54153898	92.1	n/a
Mixed	677117	1.2	14.6
Asian or Asian British:			
Indian	1053411	1.8	22.7
Pakistani	747285	1.3	16.1
Bangladeshi	283063	0.5	6.1
Other	247664	0.4	5.3
Black or Black British:			
Caribbean	565876	1.0	12.2
African	485277	0.8	10.5
Other	97585	0.2	2.1
Chinese	247403	0.4	5.3
Other	230615	0.4	5.0
All Minority Ethnic population	4635296	7.9	100
All population	58789194	100	n/a

Source: Census, April 2001, Office for National Statistics

- Ethnic group has not been routinely collected in cancer registration; incidence has to be estimated from country of birth collected on death certificates
- Reports suggest that both breast and colorectal cancers were lower in the South Asian population; however this is rapidly increasing over time[9]. More importantly, over the last 10 years, cancer rates have fallen among the rest of the population[9,10]
- It can be seen from the 2001 census data (fig. 1) that there is a lower percentage of Black African, Mixed and South Asian >= 65 years old when compared to White's and Black Caribbean. This could explain lower cancer incidence in certain minority groups
- This is altered when comparing the <16 year olds for which 55% are of mixed race when compared to 19% from the White population
- Progressive ageing of the minority ethnic population is anticipated in the future
- Populations are changing and ethnicity will become an important factor when determining cancer incidence

Figure 1: Age distribution by ethnic group, 2001/2002



Source: Census, April 2001, Office for National Statistics

Patients and Methods

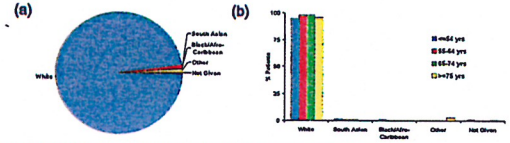
- Ethnicity data was prospectively collected in two national phase III oncology trials:
  - 1279 women with early stage breast cancer recruited between 2000-2005
    Median age = 63 years (range 35 to 90 years)
  - 2434 patients with stage II/III colorectal cancer recruited between 2002-2004
    Median age = 65 years (range 25 to 89); 64% male & 36% female
- 17 ethnic groups were collected and collapsed into 5 main groups:
  White, Black/Afro-Caribbean, South Asian, Other and Unknown/not reported
- Ethnic groups were explored across age in breast cancer and across age and gender in colorectal cancer patients

Results

Breast Trial

- 97.6% of patients recruited into the breast trial were of white origin, 1% South-Asian, 1% other, 0.4% Black/Afro-Caribbean and 0.3 were not given (fig. 2a)
- There were roughly equal numbers of white patients when split by age (fig. 2b)
- There was a distinct lack of patients >75 years in the South Asian and Black/ Afro-Caribbean groups; the Black/ Afro-Caribbean's were <=54 years

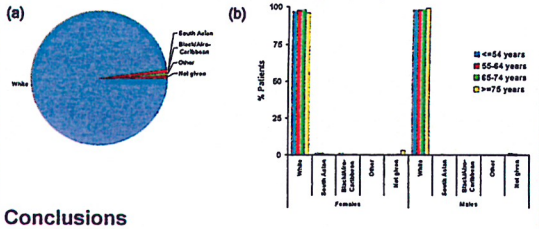
Figure 2: Breast trial patients by ethnic group and by age



Colorectal Trial

- 97.9% of patients recruited into the colorectal trial were white, 0.6% were South-Asian, 0.4% Black/Afro-Caribbean, 0.2% other and 0.9% not given (fig. 3a)
- Roughly equal numbers of white patients across age & gender (fig. 3b)
- Distinct lack of Non-White males & South Asian females >75 years

Figure 3: Colorectal patients by ethnic group, age & gender



Conclusions

- Cancer incidence is reported to be increasing within ethnic minority groups
- Ethnicity data must be prospectively collected by Cancer Intelligence Units across the UK in order to obtain a more accurate picture of cancer incidence
- The majority of patients recruited into UK oncology trials of common cancers will be White/English-speaking patients, due to the ease of obtaining full informed consent
- Strategies urgently need to be put in place in order to increase minority participation
- Research has been conducted to determine barriers to participation into trials
- A systematic review of interventions to increase recruitment and retention of ethnic minorities into clinical trials is planned; this will help identify successful interventions across all disease sites so that these can be implemented into the oncology setting

Acknowledgments

The authors would like to thank Warwick Medical School, University of Warwick, Coventry UK and Cancer Research UK who both provided generous travel bursaries; and all patients who kindly provided their ethnicity data

References

1. Mason S, Hussain-Gambles M, Leese B, Atkin K, Brown J. Representation of south Asian people in randomised clinical trials: analysis of trials' data. British Medical Journal 2003;326:1244-5
2. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials - Race-, sex-and age- based disparities. Jama-Journal of the American Medical Association 2004;291:2720-6
3. NIH Revitalization Act 1993 http://grants2.nih.gov/grants/peer/hum_anim_notice.pdf
4. Ward LC, Slegg GP, Dunn JA, Harvey P, Hallissey MT, Hyde C, Begum G, Harris S, Kerr DJ. A systematic review reporting doctors' and patients' attitudes toward participation in clinical research trials. NHS National Cancer R&D Programme Project
5. Hussain-Gambles M, Leese B, Atkin K, Brown J, Mason S, Tovey P. Involving South Asian patients in clinical trials. Health Technology Assessment 2004; 8: 42
6. Cancer Research UK, Scientific yearbook 2004/2005
7. White A. Social Focus in Brief: Ethnicity 2002. Office for National Statistics, London. [http://www.statistics.gov.uk/statbase]
8. Office for National Statistics http://www.statistics.gov.uk/pci/nugget.asp?id=273
9. Farooq S, Coleman MP. Breast cancer survival in South Asian women in England and Wales. Journal of Epidemiology and Community Health 2005;59(5):402-406
10. Smith LK, Botha JL, Benghiat A, Steward WP. Latest trends in cancer incidence among UK south Asians in Leicester. British Journal of Cancer 2003;89(1):70-73

Email: g.begum@warwick.ac.uk

WARWICK

University Hospitals Coventry and Warwickshire NHS Trust

CANCER RESEARCH UK

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)[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[2]

Figure 1

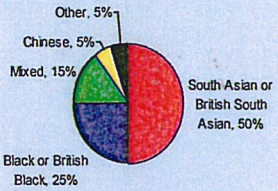
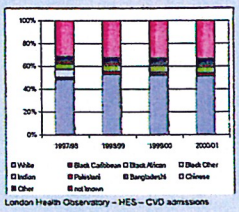


Figure 2



London Health Observatory – HES – CVD admissions

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'

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

Methods

- Literature searches were carried out using Embase, PsychIt, MEDLINE, PsycINFO, CINAHL and Google/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

	1990-1999	2000-2007	Total
Category	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) 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 31st 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 data 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[3], whilst SANGRA has a 89-91% sensitivity and 94-98% specificity[4].

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

1. Office for National Statistics http://www.statistics.gov.uk/cci/nugget.asp?id=273
2. London health observatory http://www.lho.org.uk
3. 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
4. Nanchahal K, Mangtani P, Alston M, das 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

CanEth: Improving ethnicity data collection for statistics of cancer incidence, prevalence and survival in the United Kingdom – public and professionals perceptions

Gulnaz Iqbal[1], Anil Gumber[1], Mark Johnson[2], Ala Szczepura[2], Sue Wilson[3], Janet Dunn[1]
[1]University of Warwick, UK, [2]De Montfort University, Leicester, UK, [3]University of Birmingham, UK

Background

- Accurate 'ethnicity' data essential to inform officials of incidence, prevalence and outcomes of specific diseases in population subgroups
- Some ethnic minorities associated with increased incidence of diabetes, hypertension, stroke and certain cancers
- 4.5 million people (8%) from UK in 2001 defined as being from an ethnic minority group
- Immigrants from Pakistan, India, Bangladesh combined into South Asian group (figure 1)
- South Asians largest ethnic group in the UK (50% of non-white population, 2001)
- Reports suggest breast and colorectal cancer incidence lower in South Asian population; however this is rapidly increasing over time
- UK Government initiatives in place to collect ethnicity data since 1995 but limited to hospital admissions
- Data remains incomplete and has not improved over time

Figure 1: South Asia

Objectives

1. Evaluation of health care professionals' perceptions and experiences of collecting ethnicity data in primary/secondary care
2. Evaluation of consumers' perceptions, experiences and willingness to provide ethnicity data in primary/secondary care

Methods

1.Healthcare professionals survey

- 2-page questionnaire distributed through: Minority-Ethnic-Health jiscmail list, ALLSTAT jiscmail list, National Cancer Research Network, Race-for-health Primary Care Trusts
- Questionnaire was aimed all levels of healthcare professionals involved in data collection

2.Focus groups

- Focus groups formed from existing contacts with volunteer groups and facilitators to include main minority groups in local area
- Facilitators used a topic guide specifically developed to focus on the five areas of interest:

1. General opinions on the collection of ethnicity
2. Experiences of providing ethnicity
3. Categories used in practice (examples provided)
4. How should this information be collected
5. Closing comments

1. Healthcare professionals survey results

- Perceived importance of collecting ethnicity data

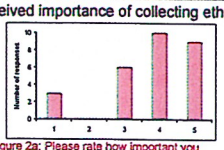


Figure 2a: Please rate how important you personally think the collection of ethnicity data is

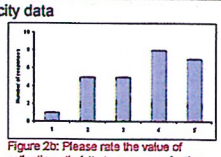
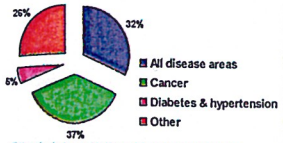


Figure 2b: Please rate the value of collecting ethnicity to your organisation

- Majority of respondents personally thought collection of ethnicity data was more important than their perception of it's value to their organisation
- May be indicative of the organisations weak policies on and lack of training provision (figures 2a and 2b)

Ethnicity data collection

Figure 3: Disease areas



Other includes: midwifery, all hospital registrations, contraception & sexual health

Table 1:Ethnicity data items collected

Data item	Yes	No	Missing
Gender	16	0	1
Race	5	11	1
Religion	11	6	0
Patient name	6	6	5
Country of birth	6	12	1
Country of origin	5	11	3
Language	10	6	3
Other	1	3	15

* Need for interpreter

- Ethnicity data routinely collected by 66% of respondents
- Most commonly collected for cancer (figure 3)
- Census, religion and language most commonly collected (table 1)
- Self- assessment most common method of data collection (n=22)

Problems collecting ethnicity data

"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." (Profiling officer, Liverpool)

"People collecting the data may not realise that they have to ask the patient" (Informatics official, London)

"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)

"Existing data collection systems are not made for it. Ethnic categories are not up to date, follow old traditional immigration routes" (Information Analyst, Luton)

Reasons for not collecting ethnicity data

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

"Not relevant to care or treatment given to patients..." (Oncology Research Sister, York)

"Our data collection is poorly resourced as it is so we have to stay entirely focused on what is clinically relevant" (Oncologist, Birmingham)

"...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 official, London)

"Not part of my job" (Radiographer, Gloucestershire)

Survey summary

- Self-assessed ethnicity is most common method of collection
- Data not collected due to a lack of 1) understanding 2) resources 3) training
- 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

Focus group results

Group	Country of Origin	Language	Males	Females	Total
1	Azad Kashmir	Mirpuri	0	5	5
2	Bangladesh	Bengali	8	0	8
3	Pakistan	Urdu	0	10	10
4	Pakistan	Urdu	8	0	8
5	India	Punjabi	2	3	5
Total			18	18	36

Selected quotes:

"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]

"Not routine; there is no need since these things don't change but once or twice is ok" [Mirpuri female]

"The information should be collected at the GP surgery as patients are already distressed in hospital" [Punjabi female]

Focus group summary

- No objection to providing data for healthcare purposes
- Explaining why data is needed and its use would increase willingness
- Ethnicity should only be collected once by GP or at first hospital visit
- There was a feeling data collected for 'statistical purposes' not used
- 'Ethnicity' information should include language, religion and country of birth to account for cultural differences

Conclusions

- Need more reporting of ethnicity data in the healthcare setting, in order to improve planning and delivery of services for ethnic minority groups
- Need training to raise awareness for patients and professionals:
  - Patients- why your doctor should know your ethnic group?
  - Professionals- how to ask/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

Contact details

Email: G.Iqbal@warwick.ac.uk for further details
This CR UK commissioned study was jointly funded by CR UK and University of Warwick

The changing age structure of the UK ethnic population, 1991 to 2001

Gulnaz Iqbal, Janet Dunn, Margaret Thorogood
Warwick Medical School, University of Warwick, UK

Background

- Population age structures (population pyramids) of Black Minority and Ethnic Groups (BMEGs) are rapidly changing within the UK
- These changes are influenced by a number of factors, the three most important being:
  1. Initial migrants are ageing and many are now >65 years old
  2. Migrants are returning to their country of birth as they become older
  3. New second/third generation UK-born ethnic minorities
- Numbers of people self-reporting as belonging to a BMEG is increasing
- The 1991 census reported over 3 million people (5.5%) of the UK population identifying themselves as belonging to a non-white ethnic group
- By the 2001 census, this had increased to 4.6 million (7.9%)
- Ageing populations have an impact on health services and health needs, especially for cancer where the majority of patients diagnosed are >65 years
- Emerging evidence indicates that certain ethnic minorities are at a higher risk of developing cancer
- This presentation explores changing population pyramids of BMEGs to estimate the likely changing demands for cancer care

Methods

Data source =Office for National Statistics/ref owen
Population pyramids created using Excel
Population pyramids exploring age, sex and country of birth across BMEGs based on the 2001 census were produced and compared to the 1991 census. The ethnic grouping for England and Wales, Scotland and Northern Ireland were combined to give 11 common groups

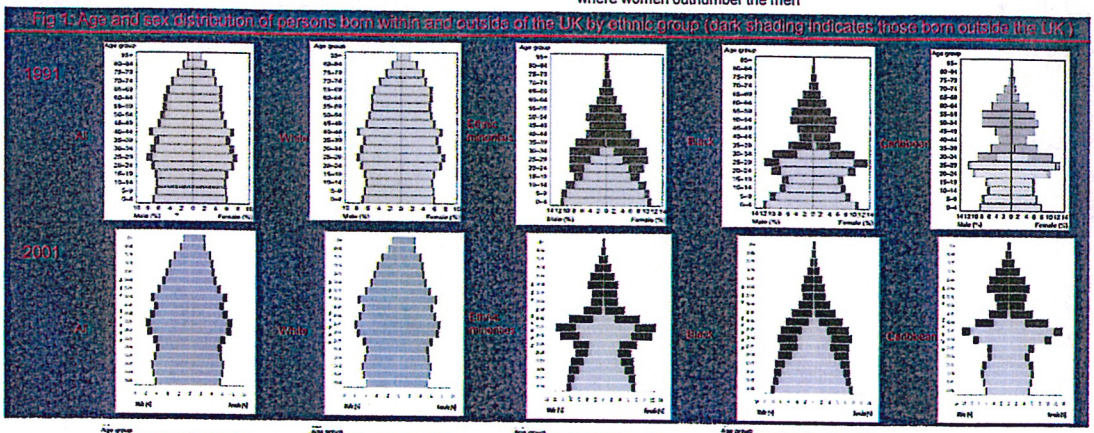
Objective

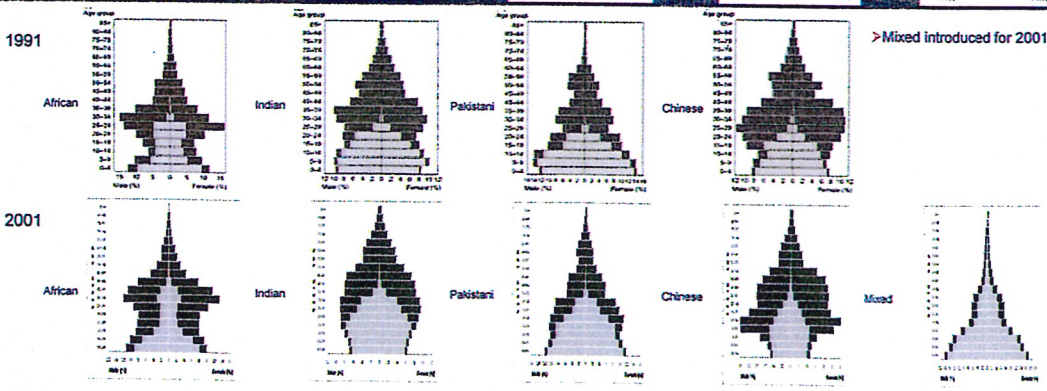
To compare the age and sex structure of the ethnic population born within and outside the UK from the 1991 to 2001 census

Results

- BMEGs have a younger age structure of BMEGs compared to the White population
- A protrusion can clearly be seen working its way through the population as a direct result of the immigration boom of the 1950s/60s
- In the elderly Bangladeshi and Pakistani groups a gender imbalance was noted. This imbalance is reversed in the Black-Caribbean population where women outnumber the men

Fig 1: Age and sex distribution of persons born within and outside of the UK by ethnic group (dark shading indicates those born outside the UK)





- Mixed introduced for 2001

Conclusions

- There is clear evidence that the population structure of BMEGs are changing and cancer services need to take account of these changes in order to continue to provide optimal services.

Contact details

Appendix 5: Published article 1: Iqbal 2009

Diversity in Health and Care 2009;6:xxx–xxx

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 Clincal 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 19961997 had incomplete data, whereas by 20012002 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, Psychinfo 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).

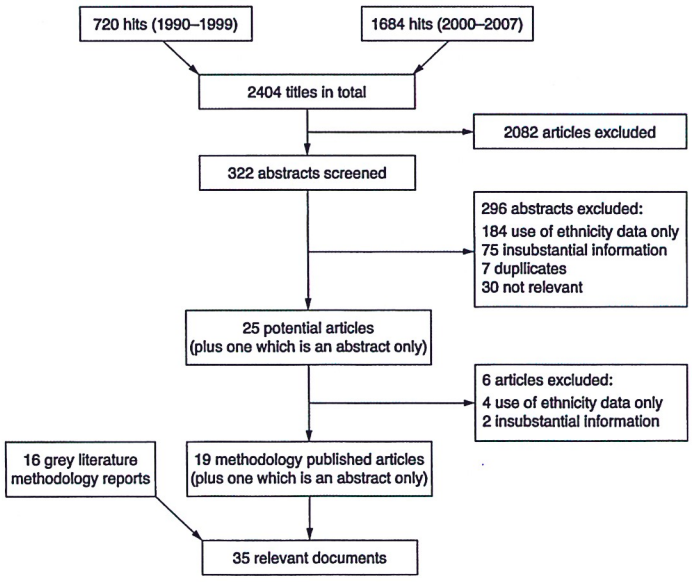


Figure 1 Ethnicity data collection and monitoring review selection process.

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 *et al*	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 intel- ligence 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, 19941997. *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

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.isdscotland.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 Clincal 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

Appendix 6: Published article 2: Iqbal 2012a

Diversity and Equality in Health and Care 2012;9:xxx–xxx

Research paper

Ethnicity data collection in the UK: the healthcare professional's perspective

Gulnaz Iqbal BSc (Hons)
Research Fellow, Warwick Clinical Trials Unit, 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

Anil Gumber PhD
Principal Research Fellow, Faculty of Health and Wellbeing, Sheffield Hallam University, Sheffield, 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

- Inequalities in health and access to healthcare among ethnic minority groups have been widely reported. However, ethnicity data collection to date has often been conducted in an ad-hoc manner, resulting in patchy data.
- The need for good-quality and complete ethnicity data has been reinforced by the Equality Act 2010, which places responsibility on authorities to tackle inequalities and target services appropriately.
- Many authors emphasise the need for improved ethnicity data collection and monitoring, but little is known about the barriers that healthcare professionals face when collecting these data.

What this paper adds

- It provides a rare glimpse of the barriers to ethnicity data collection as revealed by healthcare professionals who are not collecting the data, and a better understanding of the problems experienced by those who do collect them.
- The barriers identified here point to areas in which the development of training materials is vital.

ABSTRACT

The collection of ethnicity data has been demonstrated to be important in healthcare. However, despite recent efforts by the UK government, it remains incomplete and unvalidated. In order to be able to assess inequalities and target resources appropriately, it is essential to have complete and accurate data. This paper examines the reasons for the gaps in ethnicity data based on the perceptions and experiences of the healthcare professionals who are charged with collecting these data.

A questionnaire was used to assess perceptions of ethnicity data collection, including any barriers encountered as well as the perceived importance of collecting these data. Respondents were asked whether routine ethnicity data collection was limited to specific disease areas and approximately what proportion was

complete in these areas. There were also questions concerning preferred methods of collection (e.g. self-report). The questionnaire was completed by 30 respondents, who included healthcare managers, clinicians, nurses and other staff working in the healthcare setting. The findings confirmed that the collection of patients' ethnicity data is deemed important by the healthcare professionals, but showed that there remains uncertainty and unease as to how best to collect these data or how to explain to patients how the data will be used. The majority of healthcare professionals agreed that it was important to record patients' ethnicity, but no clear rationale was given to staff about the use of these data, and no training was provided on the best way to collect the data.

Keywords: data collection, ethnicity, healthcare professionals, perspectives

Introduction

The 2001 UK census classified 4.6 million people as belonging to a non-white ethnic group (7.9%) ,with over 50% of these being Asian or British Asian (Office for National Statistics, 2001). Reports on health inequalities and outcomes by ethnic group emphasise the necessity of overcoming barriers to make way for complete and accurate recording of ethnicity. Some ethnic groups have an increased incidence of specific diseases, such as some cancers, and experience disparities in access to both primary and secondary services (Chinegwundoh *et al*, 2006; Bowen *et al*, 2008; Sproston and Mindell, 2006; Jack *et al*, 2009, 2010, 2011; Farooq and Coleman, 2005; Atkinson *et al*, 2001; Aspinall and Jacobson, 2004). Furthermore, certain ethnic minority groups are associated with risky behaviours. For example, smoking rates are reported to be highest in Bangladeshi males, at 44%, compared with 27% in the general population (White, 2002).

Policies based on inaccurate data may lead to poor targeting of resources and services (White, 2002; London Health Observatory, 2003; Mackintosh, 2005). In this evidence-based era, the reality neatly stated by Johnson is 'that which is measured can be aimed at; that which is left unobserved can be ignored' (Johnson, 2012, pp. 3940).

UK government policy first required ethnicity data collection for Hospital Episode Statistics (HES) in 1995. However, high levels of missing data and invalid codes in the early years made the data unusable (Aspinall, 2000). Although there has been some improvement, such as the decline of 'not known' or 'not stated' codes in Finished Consultant Episodes from 23.9% in 20042005 to 8.6% in 20092010, HES data remain incomplete (HESonline, 2009). In 2004, the Quality and Outcomes Framework (QoF) began awarding one point (of a possible 1000 points) to GP practices for the recording of ethnicity for all newly registered patients. The incentive was insubstantial, and uptake was limited and was therefore later abandoned (Johnson, 2012).

In 2005, the Department of Health produced *A Practical Guide to Ethnic Monitoring in the NHS and Social Care* (Department of Health, 2005b). An ethnic monitoring tool developed by NHS Scotland was also released in 2005. The tool offered information on the 'whys' (duty and accountability) and 'hows' of ethnic monitoring (i.e. who is involved and what needs to be put in place). Training materials were available to download and modify alongside 'Training the Trainer' notes and role-play scenarios (NHS Health Scotland, 2005). However, little is known about the practical applicability and uptake of any of these guidelines.

The Equality Act 2010 reinforced the Race Relations Act of 1976 and the subsequent Amendment in 2000. This legislation made public authorities directly responsible for ensuring equity in access to healthcare and for reducing inequalities. Furthermore, public authorities are required to publish data to demonstrate their adherence to the legislation and to set clear objectives for the future. The intention is that compliance with the legislation will lead to a better understanding of the decision-making processes and make public authorities accountable for their performance (Home Office, 2010).

Staff attitudes to ethnicity data collection have been reported to be quite positive (Pringle and Rothera, 1996). For example, a survey of 16 GPs and practice managers reported that they all regarded ethnicity data collection favourably and thought it acceptable, practical and beneficial for service evaluation and targeted health promotion as well as for other purposes (Sangowawa and Bhopal, 2000). Nevertheless, ethnicity data collection is known to be patchy, particularly in primary care, where much of the research has focused upon the acceptability, feasibility, resource implications (including staff time), and limitations of computer systems, categories and coding (Kumarapeli *et al*, 2006; Pringle and Rothera, 1996).

In our first publication about this research we reported findings from a systematic literature review of ethnicity data-collection methodology in primary and secondary care (Iqbal *et al*, 2009). 'Barriers to collection' featured as one of seven themes that were

identified, and the evidence revealed healthcare professionals' perceptions to be a major obstacle to the collection of ethnicity data. Fear of causing offence to patients or encountering resistance, together with confusion about ethnicity categories and a lack of understanding of the need for ethnicity data, have also been reported as deterrents by healthcare professionals in two reports from the USA (Hasnain-Wynia *et al.* 2004; Regenstein and Sickler, 2006). Baker *et al* (2005) reported that administrative staff feared that asking for ethnicity data would alienate patients. Barriers reported by US physicians included the beliefs that collecting ethnicity data would be time consuming, would impinge on privacy and would be uncomfortable for both staff and patients, but the greatest barrier was the belief that the data had no relevance (Wynia *et al*, 2010).

In our second publication we reported the results of a series of focus groups conducted with healthy South Asian volunteers (Iqbal *et al*, 2012). The topic guide focused on perspectives and experiences of ethnicity data collection in a healthcare setting. The findings showed a somewhat linear relationship between staff comfort and patient willingness, such that the more comfortable the staff appeared to be about asking the question, the more willing the patients were to provide these data. The participants also felt that staff should be able to offer reasons for collecting the data and explanations of how the data would be used. In a US study, Baker *et al* trialled four different rationales and found that patient comfort levels were highest when quality monitoring was cited as the reason for collection (Baker *et al.* 2005).

Despite the push towards improving the completeness and reliability of ethnicity data recording, little is known about how healthcare professionals in the UK perceive the collection of these data. The aim of this paper is to explore the likely reasons for gaps in the ethnicity data by evaluating the perceptions and experiences of healthcare professionals who are tasked with collecting this information.

Methods

Cancer Research UK commissioned a project to assess ethnicity data collection for statistics relating to cancer incidence, management, mortality and survival in the UK. Ethical approval was obtained from South Birmingham Research Ethics Committee.

A survey of healthcare professionals was undertaken using a questionnaire based upon one previously developed by the Centre for Evidence in Ethnicity, Health and Diversity (CEEHD), and modified by the project working group. The modified questionnaire consisted of nine items using a mixed style of questions. Respondents were asked to rate how important they thought the collection of ethnicity data was using Likert-style items, while other questions were posed in either a closed format (no/yes/not known response options) or a tick-box format. Two open-ended questions were included, allowing respondents to provide detailed reasons for not recording ethnicity and to describe any problems encountered. The questionnaire was intended for clinicians, managers, nurses and other staff (e.g. reception staff) involved in collecting or using ethnicity data in a healthcare setting (see Table 1).

The questionnaire was distributed between March and June 2007 throughout England and Wales, via the Minority-Ethnic-Health and ALLSTAT JISCMail lists (a national academic mailing list service for academic and research communities). Questionnaires were also circulated to the 23 Race for Health primary care trust programme leads, as well as to all registered members of the Race for Health mailing list. The questionnaire was posted on the CEEHD website with a link to this placed on the NHS Evidence Ethnicity and Health (formerly the Specialist Library for Ethnicity and Health) website and sent to the National Cancer Research Network (NCRN) head office for circulation to the 24 Cancer Networks in England and Wales. A thread was created on NHS and Academic Clinical Oncology and Radiobiology Research Network (ACORRN) discussion forums. Regular weekly bulletins from the NHS forum to its members highlighted new threads. The questionnaire could be completed and returned by either post or email. In total, 14 questionnaires were completed and returned within the 4-week deadline. This was extended for a further 4 weeks (on the website links). Circulation of the questionnaire to the NCRN was repeated, but this time the questionnaire was sent electronically to each network manager, which increased the number of questionnaires returned to 30. There was a special interest in the cancer networks because the project was commissioned by Cancer Research UK.

Results

In total, 30 responses were received, coded, analysed and reported using descriptive statistics. Responses to the open questions *reasons why ethnicity data are not collected* and *problems encountered when collecting these data* are presented as direct quotations. Respondents classified themselves as *clinicians* ($n = 7$), *nurses* ($n = 5$), *managers* ($n = 5$), *information scientists* ($n = 6$) and *other* ($n = 7$), which included two radiographers, a cancer services coordinator, a patient profiling officer, a quality coordinator, a diabetes educator and a diversity manager.

Table 1 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, and racial category)

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

Unimportant				Very important
1	2	3	4	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 percentage 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:

Table 1 continued

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

Census ethnic group	No/Yes/Not known	Country of birth	No/Yes/Not known
Race	No/Yes/Not known	Country of origin	No/Yes/Not known
Religion	No/Yes/Not known	Language	No/Yes/Not known
Patient name (i.e. for use with name recognition software)	No/Yes/Not known		
Other	No/Yes/Not known		

If other, please give details

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 and 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

In total, 21 respondents (70%) attempted to routinely collect some form of ethnicity data, two (7%) did not consistently collect ethnicity data, and seven (23%) did not collect any ethnicity data. Respondents who collected ethnicity data (routinely or occasionally) did so for cancer (37%), for all disease areas (32%) or for diabetes and hypertension (5%). The majority used the recommended self-report method ($n = 12$); observer assessment was used less frequently ($n = 4$). Several respondents reported using a combination of methods (e.g. self-report and observer assessment) (see Figure 1).

Respondents who did not collect ethnicity data were asked to give their reasons for this. Their explanations included a lack of resources:

> Our data collection is poorly resourced as it is, so we have to stay entirely focused on what is clinically relevant.
>
> (Oncologist)

> 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)

Respondents stated that they were not required to collect or report ethnicity data:

> Ethnicity data is not part of the data sets that are collected.
>
> (Information manager)

In some instances, ethnicity data were only collected for specific services or when requested as part of a clinical trial:

> 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)

> 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)

Collecting ethnicity data could be problematic:

> because it involves asking the patient what they want it to be and they are not always available or willing to answer.
>
> (Informatics lead)

Staff collecting ethnicity data might not be aware of the need for self-report, and patients might refuse to answer if the options available did not match their ethnicity. Ethnicity was most commonly recorded based on the categories used in the Census (Office for National Statistics, 2001). Data about religion and language were routinely collected, but data about country of origin, race and country of birth were least likely to be collected (see Figure 2). Data systems were reported to be inadequate, and the ethnicity categories needed to be refreshed:

> Existing data collection systems are not made for it. Ethnic categories are not up to date, follow old traditional immigration routes.
>
> (Information analyst)

> Collected as part of a large data set, and some items are poorly returned.
>
> (Chair of information network)

The optional nature of ethnicity data collection was an additional factor. Patients could choose not to respond, which meant that ethnicity data would always be incomplete.

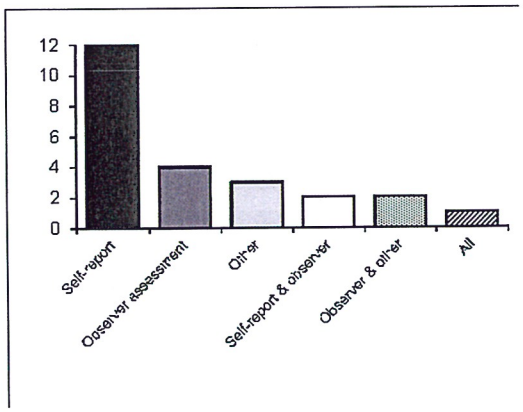


Figure 1 Method of collection.

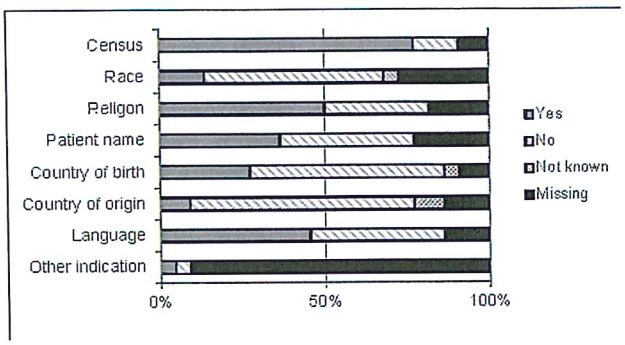


Figure 2 Routinely collected indicators.

Ethnicity data collection was considered:

> Time consuming reception time, patient time, data entry time, also language have used link workers to help patients fill in. Definite resource implications.
>
> (GP)

Patients and staff did not always understand why it was necessary:

> 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 staff's 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)

Ease of access to data was also a problem. Ethnicity data were not always recorded in an accessible place, such as the front of the patient's records or on the computer, which could mean that it was necessary to manually search for the information:

> 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)

Staff feared being challenged by patients who wanted to know the reasons for the collecting of ethnicity data, and the possibility of ensuing hostility, or causing offence:

> Patients will ask why you need to know. If they come for anonymous info [they] 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 and support services manager)

> 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)

Training large workforces was problematic:

> 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 and Human Rights, who have produced an excellent Train the Trainer pack for patient equality monitoring.
>
> (Manager)

However, the situation had changed following the development of a Train the Trainer pack by the NHS Wales Centre for Equality and Human Rights. Finally, respondents were asked to rate how important they personally thought ethnicity data collection was, and how important they perceived it to be to their organisation. Overall, respondents attached more importance to it at a personal level (69%) than at an organisational level (59%). This may be due to a lack of training provision. Only 28.5% reported that their organisation provided ethnic monitoring training, 28.5% reported that no training was provided, and the remaining 43% were not aware of any training provision. In total, 12 respondents (44%) expressed an interest in attending a workshop on ethnic data monitoring and its uses in cancer.

Discussion

These findings showed that, although individuals regarded ethnicity data collection as important, this did not mean that they went on to actually collect this

information. A number of barriers were identified, particularly that of using self-report, which is unanimously agreed to be the ideal method of collection and is recommended by many guidelines as the gold standard (Commission for Racial Equality, 2002; Department of Health, 2005b; Regenstein and Sickler, 2006). However, assessment by observation alone based upon appearance (e.g. skin colour, hair colour and/or type, dress code), despite being discouraged, was the second most commonly utilised method. Reasons given for using this approach included the avoidance of discomfort and confrontation, accompanied by a fear of causing offence to patients. This is not a new concern. An early experimental project on collecting ethnicity data in the NHS recorded a similar fear of offending patients coupled with a fear of being accused of discrimination, embarrassment when asking the questions, and concern that the questions were too sensitive as the main barriers to ethnicity data collection (Johnson *et al*, 1993). In the modern context, worries emerged about dealing with younger patients who are more likely to be born in the UK and who may wish to identify themselves as British. In addition, one respondent in our survey highlighted the difficulty of obtaining self-reported ethnicity data (as recommended by Department of Health guidelines) in situations where the patient is unwilling or simply chooses not to provide the information (Department of Health, 2005a).

Methods of data collection other than self-report are actively discouraged by the Commission for Racial Equality, and may only be used where self-report is not possible. Results of surveys conducted by the Robert Wood Johnson Foundation (a funding body established in the USA in 1972) revealed that 61% of healthcare professionals used the self-report method, whereas 25% used the observer method. Professionals considered the data to be accurate, given their knowledge of the local population, and believed that this method eliminated discomfort both for themselves and for the patients (Regenstein and Sickler, 2006).

A number of our respondents did not attempt to collect any form of ethnicity data. In many cases this stemmed from their own or their organisation's lack of awareness of the importance of the data, and the belief that it was not relevant to patient care or treatment. However, exceptions occurred if the information was required for participants in a clinical trial or for religious, dietary or communication purposes. Interviews conducted with physicians in the USA revealed that the strongest objection to collecting ethnicity and race data is the belief that it is, or should be, clinically irrelevant. Other barriers that were reported included a lack of resources, concerns about privacy, the legality of collection, and discomfort or resistance on the part of patients and staff (Hasnain-Wynia *et al*, 2010). These findings concur with the earlier results reported by Regenstein and Sickler (2006), who found that the single most important barrier to data collection is staff not knowing why it is important. However, this was not reflected in our sample, where only one participant expressed the view that it was 'not relevant to care or treatment.'

Additional barriers that were reported included difficulty in allowing staff time away from work to attend off-site training courses. However, training packages such as those developed by Lambeth Primary Care Trust, NHS Health Scotland, and the Health Research and Educational Trust (HRET) in the USA freely offer a wide range of material online, including role-play scenarios which can be used for in-house training (NHS Health Scotland, 2005; Health Research and Educational Trust, 2007; Race for Health, 2006). Weinick *et al.* (2007) have found 'train the trainer' sessions to be a viable alternative to releasing numerous staff for training in Massachusetts in the USA.

Example of good practice

Lambeth Primary Care Trust is an example of good practice where ethnicity monitoring has been relentlessly pursued. Lambeth introduced the 'Individual Patient Registration Profile' programme, which provided substantial cash injections to GP practices as an incentive to collect comprehensive patient profiling data, and also provided 1.5 days of staff training, with the half day being held at the practice. Practices were also assisted with patient profiling data collection for all patients. Mailshots of the profiling questionnaire were posted out to capture data for registered patients with free return envelopes and fully funded data entry upon return. Data were collected prospectively for all new registrations and recorded on dedicated templates provided by the programme. The resulting data have been used in a health equity audit of Stop Smoking Services and a needs assessment exercise undertaken with the Portuguese community (Race for Health, 2006).

Regenstein and Sickler (2006) have provided examples of good practice in the USA, which include the provision of ethnicity data collection training for new hospital employees as part of their induction programme. Furthermore, members of staff working in registration areas are subjected to a quality review. Managers are able to identify individuals who record a large number of 'unknown' ethnic categories or fail to record any ethnicity data, and then provide further training where necessary.

Limitations

This study was limited in terms of time and resources, and we were therefore unable to recruit a large sample or conduct large mailshots, but relied instead on links to the questionnaire posted on websites, forums, newsletters and mailing lists, and a small mailshot to all NCRN network managers. Unfortunately, this means that we are unable to calculate a response rate. In the event, we received only 30 responses despite extending the deadline for returns. With hindsight, an online questionnaire would have been easier to complete. It would have eliminated the need to print out, post or email the completed questionnaire, and might have resulted in an increased response rate. Targeted mailshots such as that to the NCRN could have been sent to other groups (e.g. individual GP practices or primary care networks), which might have yielded more questionnaire returns than using JISCMail lists. However, given the scarcity of research in this area, the responses that we did receive provide a useful insight into the perceptions and experiences of healthcare professionals today, and identify important areas for further consideration.

Conclusion

Our findings are likely to be irrelevant without a change in local and national policy. Ethnicity data collection needs to be mandated in primary care and improved in terms of quality and completeness in secondary care. Training exercises should include familiarising healthcare professionals with the Equality Act 2010, and raising awareness of the need for ethnicity data collection and how these data will be used. Methods of collection should also be included, and the importance of self-report emphasised, as well as the need for standardising of the rationale, wording of questions, response categories offered, and answers and explanations to frequently asked questions. Training may help to alleviate any anxiety felt by staff who are tasked with obtaining ethnicity data from patients. It should be emphasised that using the data we already have, irrespective of its quality and completeness, will encourage improved collection by highlighting any inadequacies. Unused data are a disincentive to healthcare professionals and patients alike (Iqbal *et al*, 2012; Fulton, 2010).

In conclusion, 'health equality is not possible without ethnic monitoring' (Fulton, 2010, p. 5). Improving ethnicity data collection requires commitment from governing bodies and agreement on what is to be collected and when. Standardised questions should be complemented by sufficiently flexible options to facilitate responses from those who do not quite fit predetermined categories. Patients need to feel assured that these data will be treated confidentially and used appropriately (Johnson, 2012; Fulton, 2010). Ethnicity data are of no value if they are not utilised to target resources and reduce inequalities (Raleigh, 2008).

A few primary care trusts have worked hard to improve ethnicity data collection, and have utilised the resulting data to help to reduce health inequalities (Race for Health, 2006; Public Health Sector, School of Health and Human Sciences, Liverpool John Moores University, 2000). However, these are isolated examples. What is needed, alongside these efforts, is a consistent message from policy makers and managers to frontline staff that collecting these data matters. Ethnicity data collection should be part of the daily routine at both primary and secondary care levels. Most importantly, we need more reporting of ethnicity data in healthcare in order to improve planning and delivery of services for members of ethnic minority groups.

ACKNOWLEDGEMENTS

The authors would like to thank all of the respondents, without whom this research would not have been possible. We would also like to thank the Advisory Board for their input into this project. This work was commissioned and funded by Cancer Research UK, and we would like to thank Vanessa Gordon-Dseagu, Catherine Foot and Ruth Yates for their support throughout the project.

REFERENCES

Aspinall P (2000) The mandatory collection of data on ethnic group of inpatients: experience of NHS trusts in England in the first reporting years. *Public Health* 114:254–9.

Aspinall PJ and Jacobson B (2004) *Ethnic Disparities in Health and Health Care: a focused review of the evidence and selected examples of good practice.* London: London Health Observatory.

Atkinson M, Clark M, Clay D *et al* (2001) *Systematic Review of Ethnicity and Health Service Access for London.* London: Regional Office, NHS Executive.

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* 20:895–900.

Bowen R, Duffy S, Ryan D *et al* (2008) Early onset of breast cancer in a group of British black women. *British Journal of Cancer* 98:277–81.

Chinegwundoh F, Enver M, Lee A *et al* (2006) Risk and presenting features of prostate cancer amongst African-Caribbean, South Asian and European men in North-east London. *BJU International* 98:1216–20.

Commission for Racial Equality (2002) *Ethnic Monitoring: a guide for public authorities.* London: Commission for Racial Equality.

Department of Health (2005a) *Department of Health Race Equality Scheme: 20052008.* London: Department of Health.

Department of Health (2005b) *Practical guide to Ethnic Monitoring in the NHS and Social Care.* London: Department of Health.

Farooq S and Coleman MP (2005) Breast cancer survival in South Asian women in England and Wales. *Journal of Epidemiology and Community Health* 59:402–6.

Fulton R (2010) *Ethnic Monitoring: is health equality possible without it?* Better Health Briefing Paper 21. London: Race Equality Foundation.

Hasnain-Wynia R, Pierce D and Pittman M (2004) *Who, When, and How: the current state of race, ethnicity, and primary language data collection in hospitals.* New York: The Commonwealth Fund.

Hasnain-Wynia R, Van Dyke K, Youdelman M *et al* (2010) Barriers to collecting patient race, ethnicity, and primary language data in physician practices: an exploratory study. *Journal of the National Medical Association* 102: 769–75.

Health Research and Educational Trust (2007) *HRET Disparities Toolkit: a toolkit for collecting race, ethnicity, and primary language information for patients.* www.hretdisparities.org

HESonline (2009) *How Good is HES Ethnic Coding and Where Do The Problems Lie?* Leeds: Health and Social Care Information Centre. www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=171

Home Office (2010) *Equality Act 2010.* London: Home Office.

Iqbal G, Gumber A, Johnson M *et al* (2009) Improving ethnicity data collection for health statistics in the UK. *Diversity in Health and Care* 6:267–85.

Iqbal G, Johnson MRD, Szczepura A *et al* (2012) UK ethnicity data collection for healthcare statistics: the South Asian perspective. *BMC Public Health* 12:243.

Jack RH, Davies EA and Moller H (2009) Breast cancer incidence, stage, treatment and survival in ethnic groups in South East England. *British Journal of Cancer* 100: 545–50.

Jack RH, Davies EA and Moller H (2010) Prostate cancer incidence, stage at diagnosis, treatment and survival in ethnic groups in South-East England. *BJU International* 105:1226–30.

Jack RH, Davies EA and Moller H (2011) Lung cancer incidence and survival in different ethnic groups in South East England. *British Journal of Cancer* 105:1049–53.

Johnson M (2012) Problems and barriers in the collection of ethnicity and migration related data. In: Ingleby D, Krasnik A, Lorant V *et al* (eds) *Health Inequalities and Risk Factors Among Migrants and Ethnic Minorities.* Antwerp: Garant.

Johnson MRD, Hardman C and Jones A (1993) *Developing the Implementation of Ethnic Monitoring: final report to the Home Office.* Warwick: EMDP and Centre for Research in Ethnic Relations, Warwick University.

Kumarapeli P, Stepaniuk R, De Lusignan S *et al* (2006) Ethnicity recording in general practice computer systems. *Journal of Public Health* 28:283–7.

Public Health Sector, School of Health and Human Sciences, Liverpool John Moores University (2000) *Ethnicity Profiling in Primary Care: the Princes Park Health Centre model.* Liverpool: Public Health Sector, School of Health and Human Sciences, Liverpool John Moores University.

London Health Observatory (2003) *Diversity Counts. Ethnic health intelligence in London: the story so far.* London: London Health Observatory.

Mackintosh M (2005) *London the world in a city: an analysis of 2001 Census results.* London: Data Management and Analysis Group, Greater London Authority.

NHS Health Scotland (2005) *Ethnic Monitoring Tool.* www.isdscotland.org/isd/files/ETHNIC%20MONITORING%20TOOL.pdf

Office for National Statistics (2001) *UK Census 2001.* www.statistics.gov.uk/census2001/census2001.asp

Pringle M and Rothera I (1996) Practicality of recording patient ethnicity in general practice: descriptive intervention study and attitude survey. *British Medical Journal* 312:1080–82.

Race for Health (2006) *Lambeth PCT Peer Review.* Liverpool: Race for Health.

Raleigh V (2008) Collection of data on ethnic origin in England. *British Medical Journal* 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.* Princeton, NJ and Washington, DC: The Robert Wood Johnson Foundation and the National Public Health and Hospital Institute.

Sangowawa O 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* 2:106–8.

Sproston K and Mindell J (eds) (2006) *Health Survey of England 2004. Volume 1. The health of minority ethnic groups.* Leeds: Health and Social Care Information Centre.

Weinick RM, Caglia JM, Friedman E *et al* (2007) Measuring racial and ethnic health care disparities in Massachusetts. *Health Affairs* 26:1293–302.

White A (2002) *Social Focus in Brief: ethnicity.* London: Office for National Statistics.

Wynia MK, Ivey SL and Hasnain-Wynia R (2010) Collection of data on patients' race and ethnic group by physician practices. *New England Journal of Medicine* 362:846–50.

CONFLICT OF INTERESTS

None.

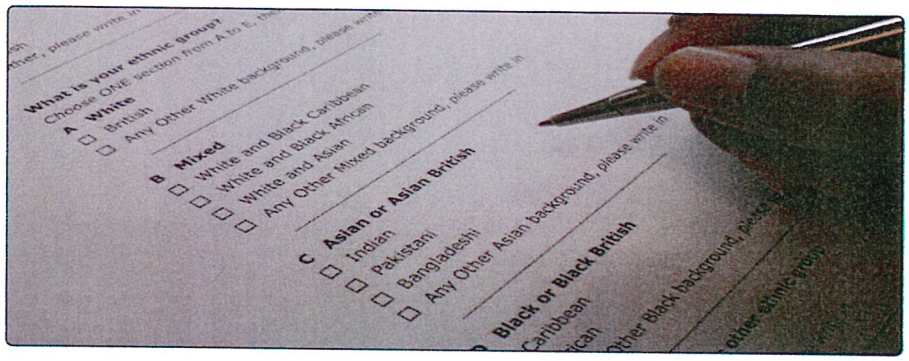
ADDRESS FOR CORRESPONDENCE

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

Received 5 July 2012
Accepted 23 September 2012

Appendix 7: Published article 3: Iqbal 2012b

BMC
Public Health



UK ethnicity data collection for healthcare statistics: the South Asian perspective

Iqbal *et al.*

BioMed Central

Iqbal *et al. BMC Public Health* 2012, **12**:243
http://www.biomedcentral.com/1471-2458/12/243 (27 March 2012)

Iqbal et al. BMC Public Health 2012, 12:243
http://www.biomedcentral.com/1471-2458/12/243

BMC Public Health

RESEARCH ARTICLE Open Access

UK ethnicity data collection for healthcare statistics: the South Asian perspective

Gulnaz Iqbal[1*], Mark RD Johnson[2,3], Ala Szczepura[2], Sue Wilson[4], Anil Gumber[2] and Janet A Dunn[1]

Abstract

Background: Ethnicity data collection has been proven to be important in health care but despite government initiatives remains incomplete and mostly un-validated in the UK. Accurate self-reported ethnicity data would enable experts to assess inequalities in health and access to services and help to ensure resources are targeted appropriately. The aim of this paper is to explore the reasons for the observed gap in ethnicity data by examining the perceptions and experiences of healthy South Asian volunteers. South Asians are the largest ethnic minority group accounting for 50% of all ethnic minorities in the UK 2001 census.

Methods: Five focus groups, conducted by trained facilitators in the native language of each group, recruited 36 South Asian volunteers from local community centres and places of worship. The topic guide focused on five key areas:1) general opinions on the collection of ethnicity, 2) experiences of providing ethnicity information, 3) categories used in practice, 4) opinions of other indicators of ethnicity e.g. language, religion and culture and 5) views on how should this information be collected. The translated transcripts were analysed using a qualitative thematic approach.

Results: The findings of this Cancer Research UK commissioned study revealed that participants felt that accurate recording of ethnicity data was important in healthcare with several stating the increased prevalence of certain diseases in minority ethnic groups as an appropriate justification to improve this data. The overwhelming majority raised no objections to providing this data when the purpose of data collection is fully explained.

Conclusions: This study confirmed that the collection of patients' ethnicity data is deemed important by potential patients but there remains uncertainty and unease as to how the data may be used. A common theme running through the focus groups was the willingness to provide these data, strongly accompanied by a desire to have more information with regard to its use.

Keywords: Ethnicity, Data collection, Perspectives, South Asians, Focus groups

Background

Over recent years there has been a drive for improved ethnicity data collection from the National Cancer Inequalities Initiative and National Cancer Intelligence Network with the main Hospital Episode Statistics (HES) data being scrutinised for completeness and validity [1,2]. Inequalities in health and access to healthcare according to ethnic group have been reported; this is of particular concern in cancer where Black, Minority and Ethnic (BME) patients have been shown to have differing rates of certain cancers compared to the general population [3-12]. A recent study showed women of African-Caribbean origin to have higher rates of breast cancer compared to the UK white population [5]. Disparities in the incidence of prostate cancer have been apparent for many years resulting in the recommended age for Black-American men to commence screening to be lowered from 50 to 40 years in the USA [13]. However, these inequalities are not restricted to cancer, disparities by ethnic group have also been observed in diabetes, where South Asians are six times more likely to develop diabetes and coronary heart disease than the general population [4]. South Asians are the largest ethnic minority group accounting for 50% of all ethnic

* Correspondence: g.iqbal@warwick.ac.uk
[1]Warwick Clinical Trials Unit, Division of Health Sciences, Warwick Medical School, The University of Warwick, Coventry, UK
Full list of author information is available at the end of the article

minorities in the UK [14]. Despite looking similar in outward appearance they differ greatly in terms of their culture, religion, language and diet.

Ethnicity data is generally known to be incomplete and of poor quality in the NHS with many still unaware of the importance of the data and its uses. Without reliable ethnicity data it is not possible to investigate differences between groups further or to develop strategies to tackle inequalities [2,15]. In 1995 it became Government policy in England and Wales to record ethnicity in Hospital Episode Statistics (HES) and in secondary care, and although there have been some great improvements such as the decline of not known/not stated codes in Finished Consultant Episodes from 23.9% in 2004-05 to 8.6% in 2009-10, HES data remains incomplete [16]. The Quality and Outcomes Framework (QoF) began awarding points (linked to financial incentives) to GP practices collecting ethnicity data on all newly registered patients in 2003. Furthermore, the collection of ethnicity data has been actively encouraged in healthcare for many years. In 2005, the Department of Health produced 'A practical guide to ethnic monitoring in the NHS and social care' which explained the relevance of data items and provided examples of good practice [17]. The drive towards the collection of complete and reliable ethnicity data stems primarily from the passing of the Race Relations (Amendment) Act (2000) which places responsibility on authorities to not only minimise inequalities but to actively promote equality. October 2010 saw the amalgamation of anti-discrimination laws to form a super Equality Act prohibiting discrimination on the grounds of nine characteristics inclusive of race and religion or belief [18].

A limited amount of research has been conducted in this area internationally, primarily in the USA. In one study of patients' attitudes towards healthcare professionals collecting data on ethnicity and race, Baker reported over half the study population to be either somewhat or very concerned (51.2%) that the data would be used to discriminate against them [19]. This proportion was significantly higher in participants of Black/African American origin compared to those of White origin (74.3% vs. 40.9% respectively).

In 1996, soon after the initiation of the mandatory ethnicity data collection in secondary care, Pringle and Rothera showed ethnicity data collection to be feasible as well as acceptable to patients and staff in the primary care setting [20]. More recently, the Information Services Division, Scotland successfully demonstrated the feasibility of collecting extra personal data (including ethnicity) for all new registrations [21]. However, there has been limited new information on how healthcare professionals and members of the public in the UK perceive ethnicity data collection despite moves to improve the completeness and reliability of ethnicity data.

The aim of this research is to explore barriers to ethnicity data collection by evaluating the perceptions and experiences of BME participants and their willingness to provide this information, investigated through a series of focus groups conducted with healthy volunteers. South Asians are the largest minority group making up 4% of the total UK population and 50% of the UK's total non-white population in 2001 [14]. Despite a similar outward appearance people originating from South Asia (most commonly India, Pakistan, and Bangladesh) are heterogeneous in terms of culture, language, religious beliefs, diet, migration history, educational attainment and social class. In order to tackle the issues of incomplete ethnicity data in health care we need to consult with these groups to not only gather their views and experiences of data provision but also on the adequacy of the fields and categories generally utilised.

This work follows on from a systematic literature review of ethnicity data collection methodology in primary and secondary care [22]. This was conducted as part of a Cancer Research UK commissioned feasibility study to improve ethnic data collection for statistics of cancer incidence, management, mortality and survival in the UK. 'Barriers to collection' was one of seven main themes identified by the systematic review and revealed healthcare professional and patient perceptions to be major obstacles to the collection of ethnicity data. Fear of causing offence to patients or encountering resistance along with confusion about ethnicity categories and a lack of understanding of the need for ethnicity data were reported as deterrents by healthcare professionals in two reports from the USA [23,24]. This paper aims to identify barriers to data collection and reports the perceptions and experiences of South Asian participants originating from Pakistan, India and Bangladesh of providing these data.

Methods

Focus groups conducted in the native language of each group was deemed the most appropriate method for this feasibility project with limited time and resources. In addition, it was felt that focus groups would allow discussion and debate between participants in what some may feel is a sensitive area.

The 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 was developed by the project team and ethical approval was obtained through South Birmingham LREC (ref: 07/Q2707/33) awarded March 2007). Focus groups were conducted by trained facilitators who recruited volunteers from local community centres and

places of worship (5-10 participants per group), where the meetings also took place. Conducting the discussions in surroundings familiar to the participants was deemed essential to create a relaxed and informal atmosphere where participants would not feel intimidated, thereby encouraging open discussion. Gender segregation was observed as per cultural custom for the Bengali and Urdu speaking participants.

Incentives were offered to encourage participation. Facilitators selected the incentive they judged would be most effective in attracting their local population. The older Bengali group were provided with refreshments including lunch after the discussion whilst the Urdu, Mirpuri and Punjabi groups received payment in the form of high street vouchers. Informed consent was taken by the facilitator where English was not the volunteers' preferred language. The facilitators used the topic guide which was specifically developed to focus on five key areas:

1. General opinions on the collection of ethnicity
2. Experiences of providing ethnicity information
3. Categories used in practice
4. Language, Religion and Culture
5. How should this information be collected?

See additional file 1 for full topic guide.

All sessions were recorded, transcribed and translated by the facilitators. Additional notes on the conduct of the groups were taken by the moderator. Each focus group discussion was subject to a quality check by an independent reviewer who listened to the recordings and validated against the translated English summary. The recordings were listened to in full and the translated transcripts provided by the facilitators were reviewed. The translated transcripts were analysed using a qualitative thematic approach which involved examining the data, comparing the accounts with one and another and identifying common themes. Themes were developed and discussed by the project working group.

Results

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, with 36 participants in total. Across groups, there was an even number of males and females. The Bengali males were the oldest group, whilst the Urdu females were the youngest (median age 63 vs. 28.5 years respectively). Data on age were not available for the Mirpuri group. A great deal of discussion in the young Urdu females group took place in English since all members had a high standard of English. For the remaining groups, the native language of the group was used in order to include all participants in the discussion. The characteristics of the total 36 volunteer convenience sample are shown in Table 1.

Table 1 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/English	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
Total			**18**	**18**	**31.5 (18-70)**	**36**

1. General opinions on the collection of ethnicity

In general, participants thought that accurate recording of ethnicity data was important. The majority were proud of their origins and were familiar with the differences between their's and other cultures, and understood the potential utility of such data in a healthcare setting. Several were also aware of the increased prevalence of certain diseases in minority ethnic groups and stated this as a reason supporting ethnicity data collection 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... However, we should be told why it is being collected when asked for it"* [Punjabi female]
- *"Sometimes certain illnesses are directly linked to our ethnicity... For example stroke or diabetes..."* [Urdu female]
- *"... 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..."* [Bengali male]

A number of participants mentioned the importance of monitoring access and uptake of services whilst others mentioned the need for collection of ethnicity for future planning. Younger participants in particular felt that it was acceptable to provide ethnicity data for health purposes but not for other reasons such as job applications:

- *"It could be alright with diseases but when you have to give this information while applying a job it would be felt like discrimination..."* [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 felt that they may be discriminated against if ethnicity was 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]

• *"Because we are all human and the same and so our ethnic origin should not interfere with the care we receive..."* [Punjabi female]

• *"It is important for government point of view but there is no importance from our point of view"* [Urdu male]

When asked whether they had any objections or worries about providing ethnicity data the majority had no objections. Several had concerns related to feelings of discomfort if the purpose of data collection was not fully explained, and expressed fears of being stereotyped. There was dissatisfaction that the appropriate ethnicity category sometimes did not appear on the form, and there was also a feeling the data would not be utilised. One participant did not think discrimination was a problem given the multi-cultural make-up of the NHS workforce:

• *"I feel uneasy sometimes and you start wondering why they ask me questions about my ethnicity"* [Urdu male]

• *"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]

• *"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]

• *"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]

2. Experiences of providing ethnicity information

In general, when asked about their experience of providing information about their ethnicity, the majority of people found it acceptable. Others expressed dissatisfaction about being asked to provide their ethnicity on repeat visits. The majority wanted some explanation as to why the data was being collected and what use it would be:

• *"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]

The main reason given for negative experiences was inappropriate codes for recording ethnicity and the fact that on several forms they would be coded as 'other', which led to feelings of frustration and insignificance:

• *"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]

• *"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]

None of the participants had an objection to providing ethnicity information in a healthcare setting. However, there was some confusion about ethnicity data collection procedures in healthcare and the need for standardisation:

• *"Sometimes they ask these questions about ethnicity and sometimes they do not so we are not sure what is the standard routine"* [Urdu male]

• *"My child was born in the same hospital yet they ask ethnic data about him whenever I took him to hospital"* [Urdu male]

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 if individuals were asked whether or not they wanted to be donors:

• *"The current ones are fine but language would be good as there are cultural differences depending on what language you speak"* [Punjabi male]

• *"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]

• *"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, most of the others in the group agreed with him]

• *"The ethnicity should not be confused with the colour of the skin"* [Urdu female]

4. Language, religion and culture

Overall, all participants were happy to disclose their religion and language as long as they did not perceive that

they were being stereotyped. The discussion on culture centred on religion being a better indicator of culture than 'ethnic group'.

- *"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]
- *"I would not hesitate to describe my language as Bengali, no reason to feel "sonkuchito" ["sense of shame"- others agreed with him]"* [Bengali male]
- *"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..."* [Urdu female]

Some Muslims did feel that they were stereotyped, especially with the heightened awareness of terrorism:

- *"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]
- *"There is always that risk in everyday life, but I guess people are far too busy with other duties to take notice"* [Mirpuri female]
- *"Yes, I feel that I am regarded as a vulnerable women because I am a non-English speaking person"* [Punjabi female]
- *"I am not Pakistani, I am a Bangladeshi. Because of my colour and appearance someone is calling me "Paki". This is stereotyping"* [Bengali male]
- *"The suspicion is that all Muslims are terrorist. This is a stereotyped view. This kinds of stereotype views should not be allowed"* [Bengali male]

Stereotyping by healthcare staff was also an issue for some participants:

- *"Walk-in centres provide independent advice but I feel my GP knows my family history so makes assumptions about me"* [Punjabi male, participant 3]

5. How should information be collected?

The Bengali focus group summarised 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. When asked about routine data collection there was a strong feeling that the data should not be collected every time as information relating to ethnicity is not likely to change very often if at all e.g. religion:

- *"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]
- *"The information should be collected at the GP surgery as patients are already distressed in hospital"* [Punjabi female, participant 1]

In summary, the majority of focus group participants had no objections to providing the data but a brief explanation of the reasons for the data collection was considered highly desirable.

Discussion

The principal findings of this Cancer Research UK commissioned feasibility study to improve ethnicity data collection for cancer statistics overwhelmingly indicates that there was no objection to providing ethnicity data for healthcare purposes in this South Asian population of focus group participants. A number of participants confidently demonstrated an understanding of differences in disease patterns by ethnic group and highlighted this as the main reason why collecting accurate ethnicity data in healthcare is of the utmost importance. There was also a consensus that ethnic group in isolation is not sufficient to capture the multi-faceted concept that is ethnicity. Many wanted additional data items such as country of birth, language and religion to be collected in order to distinguish between South Asian populations. The majority were proud of their origins and were familiar with the vast cultural differences between themselves and other South Asian communities. A small number of participants had reservations about providing the data and expressed feelings of discomfort when the purpose of the data collection and its intended use was not fully explained. Several participants expressed feelings of frustration when their ethnic group did not appear on the form and they had to tick 'other' whilst others objected to repeating the same information at every hospital visit. Most agreed GPs should collect ethnicity once and this data should be linked to hospitals. A few participants worried about disclosing their ethnicity fearing they would be labelled as terrorists, however, the majority of participants did not feel stereotyping was a problem.

This research was conducted as part of a Cancer Research UK commissioned feasibility study to improve ethnicity data collection for cancer statistics and was limited in terms of time and funding. Nevertheless, we were able to concentrate efforts on the largest minority group, South Asians made up 50% of the UK's total non-white population and 4% of the total UK population in 2001 [14]. In accordance with the cultural custom of

gender segregation the Bengali, Mirpuri and Urdu speaking groups were conducted for males and females separately, further to this, same gender facilitators were also sought for each group. Unfortunately, we were not able to find a Bengali speaking female facilitator or a male Mirpuri speaking facilitator in the timeframe of the project.

Interpretation of the findings reported here should take into account the purposeful sample and the voluntary nature of the participants, therefore this sample may be biased in favour of providing ethnicity data and results may not be generalisable to other British South Asians. To our knowledge there is little information on the perceptions of ethnicity data collection for healthcare in the UK. In contrast much has been done in the area of ethnicity data collection in the USA where the proportion of ethnic minorities is greater [25]. Despite its limitations this research has provided important messages which can be used to inform future policy and advocate the need for accurate collection of these data. The findings could also be incorporated into staff training programmes to dispel barriers to collection and address common qualms such as the fear of offending patients.

Much research into improving ethnicity data collection has been conducted in the USA where the composition of the population and healthcare systems are very different to that of the UK. The findings presented here are rich and provide a detailed picture of the views of British South Asians, building upon Pringle and Rothera's investigation into the area 15 years ago and Baker's more recent exploration of patient's attitudes to ethnicity data collection in the USA [19,20,26]. Other published work in this area includes studies reporting the feasibility of automated data linkage whereby data collected in primary care is linked through to secondary care eradicating the need for repeated collection [20,21,27].

The majority of participants considered that a brief explanation as to why the data was needed and how it would be used would increase willingness to provide ethnicity, neglecting to offering an explanation or simply telling patients it was 'routine' or 'procedure' was not deemed satisfactory. There was a strong feeling amongst some participants that data collected for 'statistical purposes' is not utilised. These findings concur with those of Pringle and Rothera and Hasnain-Wynia et al who concluded patients must be told the reason for collection and the resulting data would be used to improve the quality of services for patients [20,28]. Ultimately, evidence of data use in healthcare and government reports may be the catalyst needed to improved ethnicity data collection.

Focus group participants also stated that staff should appear comfortable when asking questions about ethnic origin. Discomfort exhibited by members of staff could make patients suspicious of the motives behind the questions and exacerbate non-compliance. Baker et al reported changes in patient comfort levels providing race and ethnicity data after hearing four different rationales, 1) quality monitoring, 2) government recommendation, 3) needs assessment and 4) personal gain. Comfort levels were shown to significantly increase ($p < 0.001$) when quality monitoring was stated as the reason for collection [26]. Exploration of similar rationale in the UK could also be informative. Well known artefacts such as the imbalance of disease burden and access to services could be incorporated into the rationale. Standardisation of the point of collection, method of collection, phrasing of questions, available responses and answers to frequently asked questions as suggested by Hasnain-Wynia et al would also be beneficial to both healthcare professionals and patients [23].

Most participants agreed general practice was the most favourable point to collect ethnicity data, where patients are less distressed and with 90% of all patient contact been with primary care there are many more opportunities to capture this information [29]. Additionally, existing patients are already acquainted with reception staff and in familiar surroundings. Collecting at the first hospital visit was also thought to be acceptable as a one-off but repeat recording at subsequent visits was not thought necessary, however repeat visits could be used as a verification point. Initiatives such as NHS electronic Summary Care Record enabling the sharing of up-to-date information will not only prove useful for healthcare professionals and reduce delays in treatment but will also ease the burden of repeatedly giving the same information for patients, ethnicity information could easily be incorporated as part of the patient demographics set [30].

Participants also discussed descriptors of ethnicity they thought to be important in healthcare and also important to distinguish between ethnic groups. Language, religion and country of birth were considered to be instrumental especially for this group of South Asians who are very different culturally in spite of having a similar outward appearance and sharing similar genetic information. One group of participants said they would describe themselves as "British Muslims" completely excluding their country of origin as they felt religious beliefs were the most significant indicator of their culture e.g. religion plays a large part in diet and consumption of alcohol and tobacco.

The findings of the focus groups reported here should be of value to healthcare professionals responsible for collecting routine ethnicity data and may help dispel some of the barriers to data collection. Common obstacles encountered by healthcare professionals are a fear

of causing offence to patients, feelings of discomfort when asking the questions and not believing the data to be of relevance. Our research shows South Asians in this sample do not mind sharing this data providing they are given a rationale and the data is used to improve services. Additionally, these findings could be used to identify data items which may be of relevance to particular local populations, additional items such as religion and diet could be added and collected as necessary. The overall aim of this work is to empower ethnicity data collection and prompt reports using this data to meet the requirements stipulated by the Race Relation (Amendment) Act 2000 and to assess whether services are currently meeting the needs of the population. This would need much more work to get right but until we have accurate and complete data on ethnicity we can't estimate rates of disease to see which services and whose needs need assessing. Simply knowing the numbers of BME patients using health services alone is not enough.

Incomplete ethnicity data has meant research to date has had little choice but to utilise methodologies such as 1) use of proxy variables where available such as Country of Birth which have distinct limitations, 2) use of name recognition software such as Nam Pehchan and SANGRA where applicability is limited to South Asians, 3) data linkage has proved useful, 4) sensitivity analyses and 5) multiple imputation or 6) conduct studies tailored to specific populations [31-37]. Landmark reports such as 'Cancer incidence and survival by major ethnic group, England, 2002-2006' produced by the National Cancer Intelligence Network are based upon incomplete data despite linking HES and national cancer registry datasets to form the National Cancer Data Repository [1]. Sensitivity analyses were conducted to assign ethnicity to the 24% of patients with missing ethnicity but crude procedures like this often lead to results that are difficult to interpret. Ryan et al reported inadequacies in both name recognition software and use of census data (ethnic distribution of area of residence) when applied to cancer registry records [38].

Downing et al opted for the more sophisticated multiple imputation in their investigation of the relationship between ethnicity and breast cancer incidence and survival as did the Office for National Statistics in their study of infant mortality for England and Wales by ethnic group [37]. However, multiple imputation is based upon untestable assumptions, in cases where ethnicity data is not missing at random multiple imputation is inappropriate e.g. missing data is concentrated in certain ethnic groups.

Further research is needed into the perceptions and experiences of ethnicity data collection in a broader range of UK ethnic groups e.g. Black Caribbeans, Black Africans, Chinese, Whites and particularly those of the rapidly growing mixed group for whom the question of ethnic group is particularly tricky.

Conclusion

It is recognised that ethnicity data collection in the UK has historically been of poor quality. Comprehensive and validated ethnicity data collection is essential if we are to reduce inequalities in health and access to healthcare services. In order to improve ethnicity data collection, the provision of training is fundamental in order to increase awareness and promote the importance and utility of recording ethnicity data for all staff that collect/use the data. Ideally, ethnicity should only be collected once by GP or at first hospital visit and linked through healthcare databases and verified at subsequent points of contact. Data collection should be extended to collect additional items such as language, religion and country of origin/birth to account for cultural differences. Only once we have complete and validated ethnicity data can we know the true extent of disparities in healthcare and devise appropriate strategies to combat them. Reducing health inequalities and tailoring current services to meet the needs of BME groups wholly depend upon having accurate and complete ethnicity, without this information we will remain blind to the size and depth of the problem, as a consequence patients with no data will inevitably be left behind.

Additional material

Additional file 1: Focus group topic guide.

Acknowledgements

The authors would like to thank all the focus group facilitators and participants without whom this research would not have been possible. We would also like to thank the Advisory Board for their input into this project and Sarah Damery for her comments on the manuscript. 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. This project benefited from facilities funded through *Birmingham Science City Translational Medicine Clinical Research and Infrastructure Trials Platform*, with support from Advantage West Midlands (AWM).

Author details

[1]Warwick Clinical Trials Unit, Division of Health Sciences, Warwick Medical School, The University of Warwick, Coventry, UK. [2]Centre for Evidence in Ethnicity Health and Diversity, University of Warwick, Coventry, UK. [3]Mary Seacole Research Centre, De Montfort University, Leicester, UK. [4]Primary Care Clinical Sciences, University of Birmingham, Birmingham, UK.

Authors' contributions

GI was the main researcher and contributed to the design and co-ordination of this study in addition to conducting the fieldwork, analysis and drafted the manuscript. MRDJ, AS, SW and AG contributed to the design and co-ordination of this project, interpretation of the data and revision of the manuscript. JAD was the Chief Investigator of the study and contributed to

the design and co-ordination of this project, interpretation of the data and revision of the manuscript. All authors have read and approved the manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 15 December 2011 Accepted: 27 March 2012
Published: 27 March 2012

References

1. National Cancer Intelligence Network: **Cancer incidence and survival by major ethnic group, England, 2002-2006.** *Book Cancer incidence and survival by major ethnic group, England, 2002-2006 (Editor ed.). City* 2009.
2. Jack RH, Linklater KM, Hofman D, Fitzpatrick J, Moller H: **Ethnicity coding in a regional cancer registry and in Hospital Episode Statistics.** *BMC Public Health* 2006, 6:8.
3. Chinegwundoh F, Enver M, Lee A, Nargund V, Oliver T, Ben-Shlomo Y: **Risk and presenting features of prostate cancer amongst African-Caribbean, South Asian and European men in North-east London.** *BJU International* 2006, 98:1216-1220.
4. Aspinall PJ, Jacobson B: **Ethnic disparities in health and health care: A focused review of the evidence and selected examples of good practice.** *Book Ethnic disparities in health and health care: A focused review of the evidence and selected examples of good practice (Editor ed.^eds.). City* 2004.
5. Bowen R, Duffy S, Ryan D, Hart I, Jones J: **Early onset of breast cancer in a group of British black women.** *British Journal of Cancer* 2008, 98:277-281.
6. The Information Centre: **Health survey of England 2004, Health of ethnic minorities.** *Book Health survey of England 2004, Health of ethnic minorities (Editor ed.^eds.). City* 2006.
7. Jack RH, Davies EA, Moller H: **Lung cancer incidence and survival in different ethnic groups in South East England.** *British Journal of Cancer* 2011, 105:1049-1053.
8. Jack RH, Davies EA, Moller H: **Prostate cancer incidence, stage at diagnosis, treatment and survival in ethnic groups in South-East England.** *BJU International* 2010, 105:1226-1230.
9. Jack RH, Davies EA, Moller H: **Breast cancer incidence, stage, treatment and survival in ethnic groups in South East England.** *British Journal of Cancer* 2009, 100:545-550.
10. Jack RH, Davies EA, Moller H: **Testis and prostate cancer incidence in ethnic groups in South East England.** *Int J Androl* 2007, 30:215-220.
11. Day M, Poole J, Bennett JA, Peake MD: **Changes in lung cancer incidence in South Asians in Leicester, 1990-2005.** *J Public Health* 2010, 32:230-235.
12. Farooq S, Coleman MP: **Breast cancer survival in South Asian women in England and Wales.** *Journal of Epidemiology and Community Health* 2005, 59:402-406.
13. Platz EA, Giovannucci E: **Prostate cancer.** In *Cancer epidemiology and preventio*.. 3 edition. Edited by: Schottenfeld D, Fraumeni JF Jr. New York: Oxford University Press; 2006:.
14. **UK Census 2001.** [http://www.statistics.gov.uk/census2001/census2001.asp].
15. **How good is HES ethnicity coding and where do the problems lie?.** [http://www.hesonline.nhs.uk].
16. HES online: **How good is HES ethnic coding and where do the problems lie?** *Book How good is HES ethnic coding and where do the problems lie? (Editor ed.^eds.). City* 2011.
17. Department of Health: **A practical guide to ethnic monitoring in the NHS and social care.** *Book A practical guide to ethnic monitoring in the NHS and social care (Editor ed.^eds.). City* 2005.
18. Home Office: **Equality Act 2010.** *Book Equality Act 2010 (Editor ed.^eds.). City* 2010.
19. Baker DW, Cameron KA, Feinglass J, Georgas P, Foster S, Pierce D, Thompson JA, Hasnain-Wynia R: **Patients' attitudes toward health care providers collecting information about their race and ethnicity.** *J Gen Intern Med* 2005, 20:895-900.
20. Pringle M, Rothera I: **Practicality of recording patient ethnicity in general practice: Descriptive intervention study and attitude survey.** *British Medical Journal* 1996, 312:1080-1082.
21. Information Services Division Scotland: **Patient profiling pilot project.** *Book Patient profiling pilot project (Editor ed.^eds.). City* 2010.
22. Iqbal G, Gumber A, Johnson M, Szczepura A, Wilson S, Dunn J: **Improving ethnicity data collection for health statistics.** *Diversity in Health and Care* 2009, 6.
23. Hasnain-Wynia R, Pierce D, Pittman M: **Who, when, and how: The current state of race, ethnicity, and primary language data collection in hospitals.** *Book Who, when, and how: The current state of race, ethnicity, and primary language data collection in hospitals (Editor ed.^eds.)* City: Health Research and Educational Trust and The Commonwealth Fund; 2004.
24. Regenstein M, Sickler D: **Race, ethnicity, and language of patients. Hospital practices regarding collection of information to address disparities in health care.** *Book Race, ethnicity, and language of patients. Hospital practices regarding collection of information to address disparities in health care (Editor ed.^eds.)* City: The Robert Wood Johnson Foundation and National Public Health and Hospital Institute; 2006.
25. **Table 6. Resident Population by Sex, Race, and Hispanic Origin Status.** [http://www.census.gov/compendia/statab/2012/tables/12s0006.pdf].
26. Baker DW, Hasnain-Wynia R, Kandula NR, Thompson JA, Brown ER: **Attitudes toward health care providers, collecting information about patients' race, ethnicity, and language.** *Med Care* 2007, 45:1034-1042.
27. Olatokunbo S, Bhopal R: **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, 2:106-108.
28. Hasnain-Wynia R, Taylor-Clark K, Anise A: **Collecting Race, Ethnicity, and Language Data to Identify and Reduce Health Disparities: Perceptions of Health Plan Enrollees.** *Med Care Res Rev* 2011, 68:367-381.
29. Raleigh V: **Collection of data on ethnic origin in England.** *British Medical Journal* 2008, 337:a1107, doi: 10.1136/bmj.a1107.
30. **Summary Care Records.** [http://www.nhscarerecords.nhs.uk/index.html].
31. Wild SH, Fischbacher C, Brock A, Griffiths C, Bhopal R: **Mortality from all causes and circulatory disease by country of birth in England and Wales 2001-2003.** *J Public Health* 2007, 29:191-198.
32. Wild SH, Fischbacher CM, Brock A, Griffiths C, Bhopal R: **Mortality from all cancers and lung, colorectal, breast and prostate cancer by country of birth in England and Wales, 2001-2003.** *British Journal of Cancer* 2006, 94:1079-1085.
33. Cummins C, Winter H, Cheng KK, Maric R, Silcocks P, Varghese C: **An assessment of the Nam Pehchan computer program for the identification of names of south Asian ethnic origin.** *J Public Health Med* 1999, 21:401-406.
34. Nanchahal K, Mangtani P, Alston M, Silva ID: **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, 23:278-285.
35. Ali R, Barnes I, Kan SW, Beral V: **Cancer incidence in British Indians and British whites in Leicester, 2001-2006.** *British Journal of Cancer* 2010, 103:143-148.
36. Fischbacher C, Bhopal R, Povey C, Steiner M, Chalmers J, Mueller G, Jamieson J, Knowles D: **Record linked retrospective cohort study of 4.6 million people exploring ethnic variations in disease: myocardial infarction in South Asians.** *BMC Public Health* 2007, 7:142, doi: 10.1186/1471-2458-7-142.
37. Downing A, West RM, Gilthorpe MS, Lawrence G, Forman D: **Using routinely collected health data to investigate the association between ethnicity and breast cancer incidence and survival: what is the impact of missing data and multiple ethnicities?** *Ethn Health* 2011, 16:201-212.
38. Ryan R, Vernon S, Lawrence G, Wilson S: **Use of name recognition software, census data and multiple imputation to predict missing data on ethnicity: application to cancer registry records.** *BMC Medical Informatics and Decision Making* 2012, **12, Article number: Provisional PDF. doi: 10.1186/1472-6947-12-3.**

Pre-publication history
The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2458/12/243/prepub

doi:10.1186/1471-2458-12-243
Cite this article as: Iqbal *et al.*: UK ethnicity data collection for healthcare statistics: the South Asian perspective. *BMC Public Health* 2012 12:243.