

Ethnic minorities in UK science: a bibliometric approach.

Berenika M. Webster

Department of Information Science,
City University,
London EC1V 0HB;
B.M.Webster@city.ac.uk

Abstract

This paper seeks to establish the degree of presence and potential impact of ethnic minority scientists on UK science. It does this by examining their published outputs (present in SCI). Ethnic minority researchers were identified and classed within seven distinct groups based on their surname. A fractional count of their papers was used to establish which fields they research into, what types of research they conduct (from basic to applied) and what is the potential impact of their work (citations to journal). Key findings suggest that in the last 20 years the proportion of ethnic-authored papers almost doubled, Chinese were the group which contributed most papers, ethnic minorities are best represented in engineering and technology, physics and chemistry and they tend to publish in applied journals with potential impact factor lower than that of non-ethnic authors.

1. Introduction

Ethnic minorities constitute an increasingly important element of the British population. Their numbers almost doubled in the last 20 years, to reach 4 million in 2000 (7.1% of the UK population) (Scott, Pearce and Goldblatt, 2001). The oldest immigrant groups in the UK are Black Caribbean and Indian – arriving in the UK from the late 1940s and throughout the 1950s and 1960s. Chinese people arrived in the UK throughout the 1980s, Bangladeshis in the early 1980s and Black Africans in the late 1980s and 1990s. In 1991 only 1% of all marriages were between partners of different ethnic origins indicating that the groups tend to stay separate (Commission for Racial Equality, 1999). This is an important factor in this research project, as its methodology relies on an analysis of surnames (ethnic vs. non-ethnic).

There is an increasing pool of evidence suggesting that this rapidly growing group is being excluded from participation in UK scientific life. An editorial in *New Scientist* pointed to the lack of black Nobel Prize winners, Fellows of the Royal Society or members of government scientific panels (*New Scientist*, 9 March 2002). It was demonstrated that ethnic minorities are being discriminated against in the process of admissions to medical schools in the UK (McManus, 1998). In academic employment in the UK ethnic minorities are less likely to be in senior posts, including professors, (Carter, Fenton and Modood, 1999) and black academics are one third as likely and Asian academics are half as likely as whites to be high earners (with a salary of £35K or more) (Association of University Teachers, 2000). Moreover, the Higher Education Funding Council for England (HEFCE) noted that ethnic minorities were underrepresented in the 1996 Research Assessment Exercise returns and called for further investigation into the issue (HEFCE, 2000).

This study is an attempt to employ bibliometric methods to assess the participation of ethnic minorities in British science. This will be done through an examination of the characteristics of publications of several groups (see: Table 1).

Table 1 Ethnic groups in the UK identified for the analysis

Group	Code	Census compatible?
Black Africans	BA	Yes
Chinese	CHI	Yes
Indians	IND	Yes
Iranians	IRA	No
North Africans and Middle Eastern peoples	NA	No
Pakistanis and Bangladeshis	P-B	Yes
Sri Lankans	SRL	No

These seven groups were selected for two reasons. Firstly, they constitute the biggest immigrant groups within the UK. Secondly, their surnames are distinct and, by and large, unique to the groupings identified above, thus relatively easy to classify. Four of these groupings also correspond with the standard classifications used in census and ethnic monitoring in the UK (see: Tab. 1). Black Caribbeans are a large and well-established immigrant group in the UK, however its members cannot be identified easily using surnames, thus they were eliminated from the analysis.

2 Definitions

It is essential to define terms and concepts that constitute the base of this study. Such terms as ethnicity and ethnic minority had to be conceptualised. It was also important to establish if a surname could be used as a surrogate for ethnicity.

An “ethnic unit” was defined as: “a population whose members believe that in some sense they share common descent and a common cultural heritage or tradition and who are so regarded by others” (Smith, 1986, p. 192). In Britain, the term ethnic minority is “widely understood to denote a category of people whose recent origins lie in the countries of the New Commonwealth and Pakistan; in other words, in former British colonies in the Indian subcontinent, the Caribbean, Africa, and sometimes the so-called Far East” (Mason, 1995, p. 15).

For the purpose of this study a surname was used to denote a person’s ethnic origin.¹ This method has been successfully used in public health studies in the USA (Lauderdale, Kestenbaum, 2000), Canada; (Sheth et al., 1997), the UK (Nicoll, Bassettand and Ulijaszek, 1986; Harding, Dews and Simpson, 1999) and Germany (Razum, Zeeb and Akgun, 2001) where surnames (sometimes in conjunction with a given name, place of birth or religion) were used to determine ethnicity of individuals. The positive results of these studies gave the author of this paper confidence that surname analysis may be a valid method in assessing peoples’ ethnicity.

3 Method

The Science Citation Index (SCI) was the main source of data in this study. Firstly, lists of surnames representing different ethnic groups were created, drawn from the SCI. For instance, to create a list of Chinese surnames a large batch of papers originating from Chinese institutions was downloaded from the SCI. From these a list of surnames was created and stripped of obviously western surnames and then passed on to a Chinese expert whose task was to exclude all remaining non-Chinese surnames. This process was repeated for all groups of surnames studied in this project. This method allowed nearly 15,000 unique surnames to be identified. Table 2 shows ethnic surnames which occur most frequently in the SCI.

¹ Surname analysis was used before in bibliometric studies to assess the presence and impact of women in science. This method can be used for populations in which male and female surnames assume different forms (see: Lewison, 2001 and Webster, 2001).

Table 2 Frequently occurring ethnic names in SCI for the selected ethnic groups.

Black Africa	Iran	China	Pakistan and Bangladesh	Sri Lanka	India	North Africa
Adebayo	Ensafi	Chan	Ahmed	Desilva	Bhatt	Abdula
Bii	Firouzabadi	Cheung	Ali	Fernando	Chander	Amin
Binta	Ghassemzadeh	Li	Begum	Gunatilake	Desai	Habib
Eke	Hajipour	Liu	Bibi	Kottegoda	Gupta	Haddad
Ekesi	Heravi	Wong	Hussain	Kumara	Kumar	Hassain
Ikerra	Iranpoor	Wan	Iqbal	Perera	Lal	Ibrahim
Kuku	Mallakpour	Xu	Khan	Ramasamy	Raje	Karim
Miah	Mohammad	Yau	Malik	Ratnasooriya	Shan	Latif
Mukaro	Shamsipur	Zhang	Naqvi	Seneviratne	Sharma	Suliman
Onigbogi	Yavari	Zhu	Siddiqui	Tennakone	Patel	Yusuf

To obtain British scientific publications, all British-authored papers were downloaded from the SCI for years 1981, 1986, 1991, 1996 and 2001 (disk years rather than publication years were used in constructing the data set). Then, the file was stripped of all papers which contained addresses outside England, Wales, Scotland or Northern Ireland (i.e. UK papers co-authored with foreign researchers). The file of UK-only papers contained 160,401 scientific papers, notes and reviews (29,119 in 1981; 31,763 in 1983; 31,939 in 1991 and 35,078 in 1996 and 32,502 in 2001). Papers published by British scientists in collaboration with foreign authors were excluded from the study as it is impossible in the SCI-derived data sets to assign an individual author to an address and thus identify which author comes from which country or institution.

It is important to note that the findings based on an analysis of UK-only papers may not necessarily be generalised to the population of all UK papers (including these with foreign co-authors). In 1981, 87% of all UK papers were domestic only whilst by 2001 this dropped to 62%. In this time we observe a drop in the proportion of non-ethnic output and a corresponding growth in the share of ethnic output (from 3.7% in 1981 to 7.2% in 2001). Because we know nothing about the internationally co-authored papers, it is impossible to exclude the possibility that the increase in ethnic output in domestic-only papers may be partly an artefact of an increased international co-operation of non-ethnic writers (but not ethnic ones). Research demonstrated (Webster, 2001) that women tend to cooperate less internationally than men do and the same could be true of ethnic authors.

To establish the volume of ethnic output, each paper was checked for the presence of an ethnic surname and a fractional value was assigned to each ethnic category present in the paper (e.g. if a paper was authored by Smith and Chen, half credit was given to the “non-ethnic” category and half credit to the Chinese category). Also, if a surname was common to two or more ethnic groupings, the fraction of a credit was assigned to each grouping based on its size. For instance, if a surname was common to Sri Lanka and India, the latter group (a bigger one in terms of authored papers) was assigned a bigger fraction of a credit than the Sri Lankan group since the probability that the author came from India rather than Sri Lanka was higher. Fractional counting of multiauthored publications seems to be a preferred method in bibliometrics (Van Hooydonk, 1997). Credit is assigned equally (and fractionally) to each co-author, as it is impossible to assess the effort of each author (i.e. we cannot rely on the sequence of authorship), while assigning integer credit to each co-author would lead to inflating the numbers of papers in some disciplines but not others (i.e. these with high instances of multiple authorship). All analyses in this study used fractional counts of papers.

To learn more about the characteristics of ethnic research, all papers were classified according to three criteria. It is important to note that all categories assigned relate to a journal title rather than an individual paper published in that journal. This means that all papers published in one journal are believed to share the same characteristics. Firstly, the papers were assigned a broad subject area (biology, biomedical

research, clinical research, chemistry, engineering & technology, earth & space, mathematics and physics). Secondly, each paper was classified by research level (RL): from basic scientific research (RL 4) through applied research (RL3), engineering-technological research (RL2) to applied technology (RL1). Subject area classification and research level categories were defined and assigned to journals by the CHI Research Inc. in the USA (Narin and Hamilton, 1986), based on expert opinion and journal to journal citation patterns. Thirdly, the papers were assigned a potential impact factor value (PIC) – a value from 1 to 4 based on the ratio between numbers of papers published and citations received in the year of publication and four subsequent years.² Category PIC 1 includes journals with 6 or fewer citations, PIC 2 – greater than 6 to 11, PIC 3 – greater than 11 to 20 and PIC 4 – greater than 20 citations.

4. Results

4.1 Population, employment and published outputs

During the five years under investigation there were 216,508 papers with UK authorship in the SCI. Over 74% of these were written without foreign collaboration, however the ratio of domestic-only and internationally-co-authored papers is shifting (from 87% of domestic-only papers in 1981 to only 62% in 2001) as international collaboration plays an increasingly significant part in UK science.

Out of 160,401 UK-only papers, 7,257 were written by ethnic minority researchers identified through surname analysis. This constitutes 4.5% of all UK-only papers. The percentage of ethnic authored papers increased with time: from 3.8% in 1981 to 7.2% in 2001. In mid-1990s for the first time the percentage of publication outputs exceeded the percentage of population for ethnic minority groups in the UK. Figure 1 illustrates the relationship between population and output levels of the ethnic minorities in the UK. Both curves lie on the same level, with the only discrepancy from early to mid-1980s when the population grew at faster rate than the published output. This may have been associated with a large influx of immigrants from Bangladesh and Africa at that time.

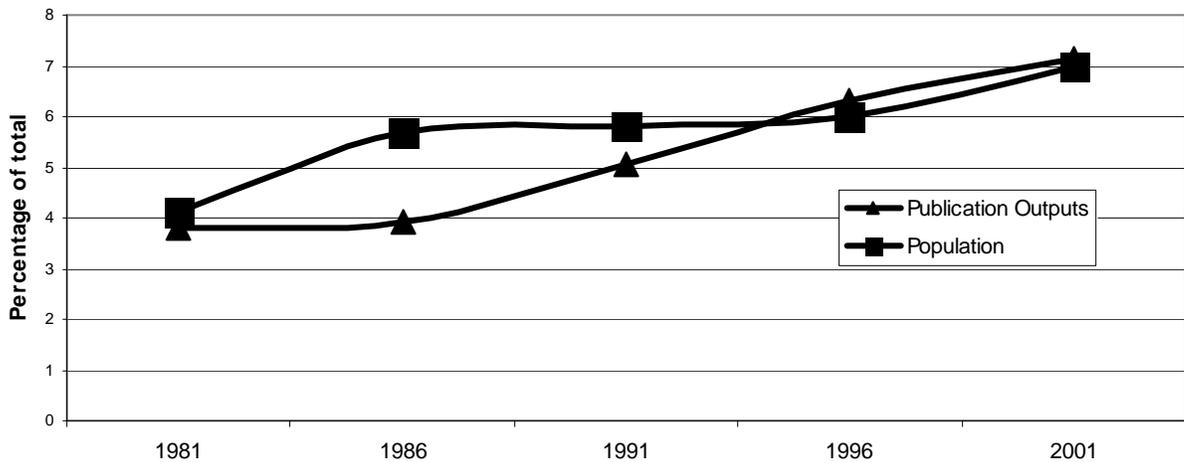


Figure 1. UK ethnic population (census and Labour Force Survey data) and publication outputs SCI, 1981-2001

The ethnic participation in British science varies across different ethnic groups (see: Fig. 2). The Chinese seem to be represented the best in relation to their overall population. In mid-1990s they constituted around 0.3% of the UK population, 1.7% of higher education staff (2.3% of staff in science disciplines) but their share in UK science output was over 2.5%. Indians, on the other hand, showed a rather balanced distribution: percentages of population, employment and published outputs were similar, with publication outputs only slightly higher than population and employment data. Black Africans seem to fare worse than

² The PIC classification comes from the Institute of Scientific Information.

other groups: despite constituting nearly 0.6% of the UK population and 0.7% academic staff in science disciplines; they only contributed around 0.1% of publication outputs. The Bangladeshis and the Pakistanis still fare differently: they constitute 1.5% of the UK population but are poorly represented in academic employment and yet produced slightly more papers than expected.

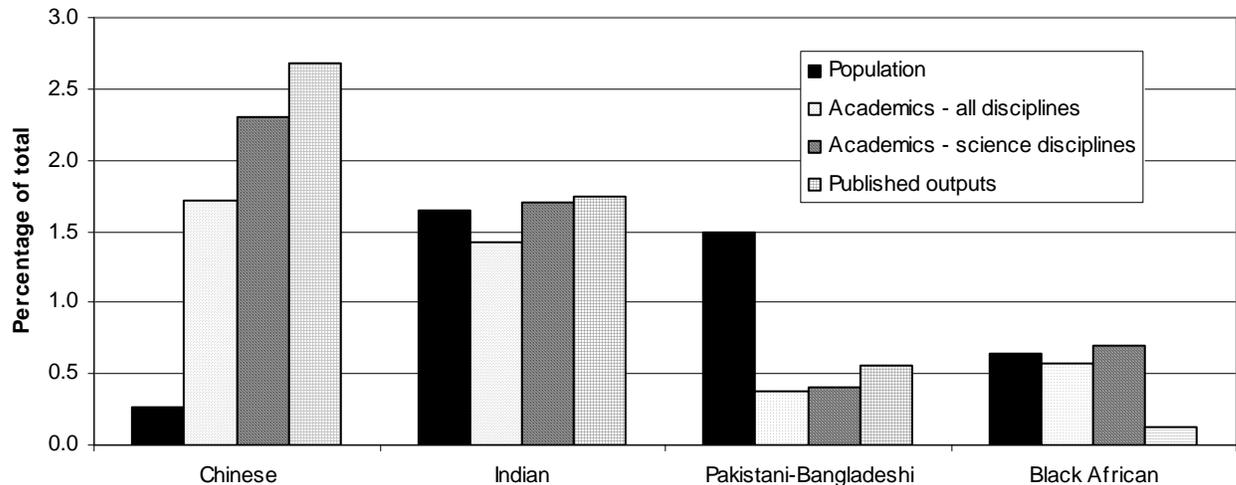


Figure 2. Mid-1990s presence of ethnic minorities in UK population (LFS), higher education employment (HESA) and publication outputs (SCI)

These differences may arise from the different status of these groups in UK society. For instance, Indians are a well-established group in the UK, with a rather mature population (their average age is considerably higher than that of other ethnic groups studied here)³ mostly born in the UK. On the other hand, Black Africans constitute a newer wave of immigrants (with large numbers born outside UK)⁴. Lower participation in higher education employment amongst Pakistanis and Bangladeshis may be because they have relatively fewer people in the age range from 20 to 64 (30% and 25% respectively compared to 42% for Indians and 45% for Chinese) (Scott, Pearce and Goldblatt, 2001) and they arrived in the UK later than Indians.

Another factor affecting the differences between groups may be variations in the levels of educational attainment amongst them. Black Africans, Pakistanis and Bangladeshis in the age group of 25 to 45 have relatively the smallest numbers of people with higher education degrees (8%, 4% and 5% respectively) while Indians and Chinese have the highest (11% and 12%).⁵

The discrepancy of population and publication counts for the Chinese is very striking. While their proportions grew little in the population surveys in the UK, the proportion of their published outputs increased dramatically. Figure 3 shows the changes in population and publication outputs between 1981 and 2001 for four major groups. Other groups grew faster, in terms of population change, than in terms of the volume of published outputs. In the case of Black Africans, the published output actually dropped (from 0.35% of all UK output to 0.25%) despite population growth from 0.15% to 0.77% between 1981 and 2001.

³ The average age for different groups is as follows: whites – 38, Indians – 30, Chinese – 29, Africans – 25 and Pakistanis and Bangladeshis – 22 (see: Jones, 1993, p. 13).

⁴ For instance in 1991 census data 42% of Indians and only 36% of Black Africans were born in the UK, while 2000 Labour Force Survey data lists 43% of Indians and 35% of Black Africans (see: Rees and Philips, 1996).

⁵ The percentage in the equivalent age group amongst whites is 8% (see: Jones, 1993, p. 46-47).

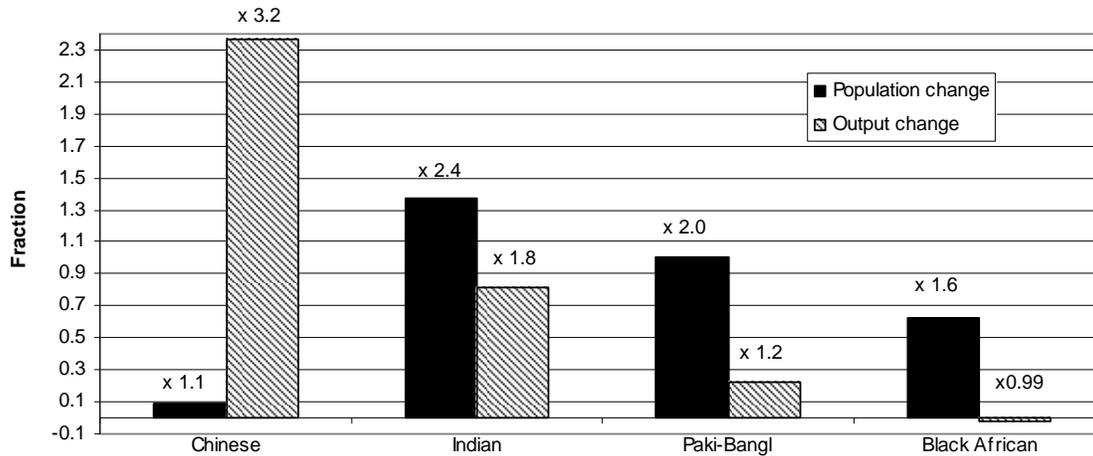


Figure 3. Percentage change in population and published outputs, 1981 and 2001

Figure 4 shows the changes in the proportion of published outputs amongst different ethnic minority groups. The Chinese are now clearly dominant, with the biggest growth (from 180.2 papers in 1981 to 969.1 papers in 2001), and are closely followed by the Indians (from 435 papers in 1981 to 750.9 in 2001). These two groups constitute around 60% of ethnic minorities employed in science in British higher education but produce nearly 70% of all ethnic published outputs. Another group showing an increase in the proportion of published outputs is Pakistanis and Bangladeshis (from 115.8 papers in 1981 to 200.7 in 2001). Other groups' contributions remain mostly unchanged over time.

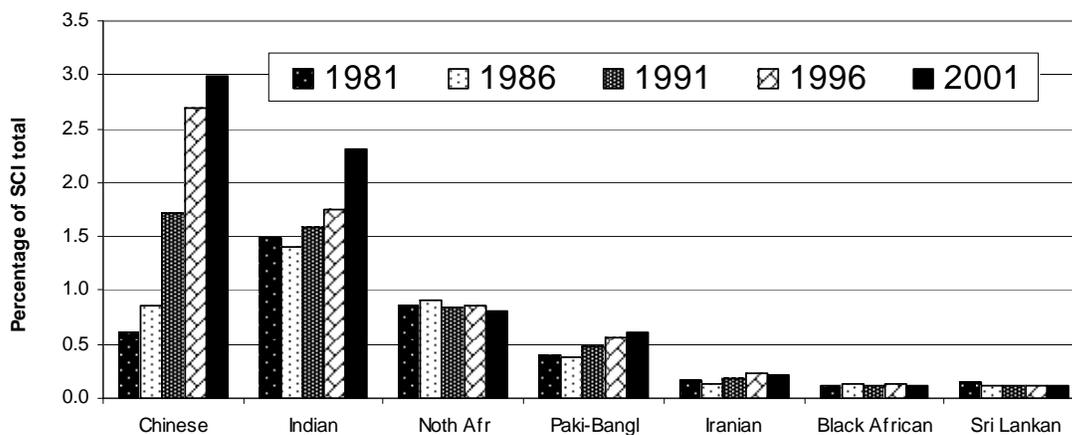


Figure 4 . Percentage of all fields papers in SCI in 1981, 1986, 1991, 1996 and 2001, by ethnic group category

4.2 Subject fields

There is a big difference in the participation of ethnic minority scholars between different scientific fields. Ethnic minorities make the largest contributions, in terms of the volume of published outputs, to the fields of engineering and technology (nearly 10% across studied years), physics (nearly 7%), chemistry and maths (over 5.5% each) and clinical medicine (slightly over 5%). The Chinese dominate all but two of the eight scientific disciplines: only in clinical medicine and biomedical research are they surpassed by Indians. Figure 5 shows the mean percentage of ethnic papers in eight fields and the contribution of different ethnic groups in these fields.

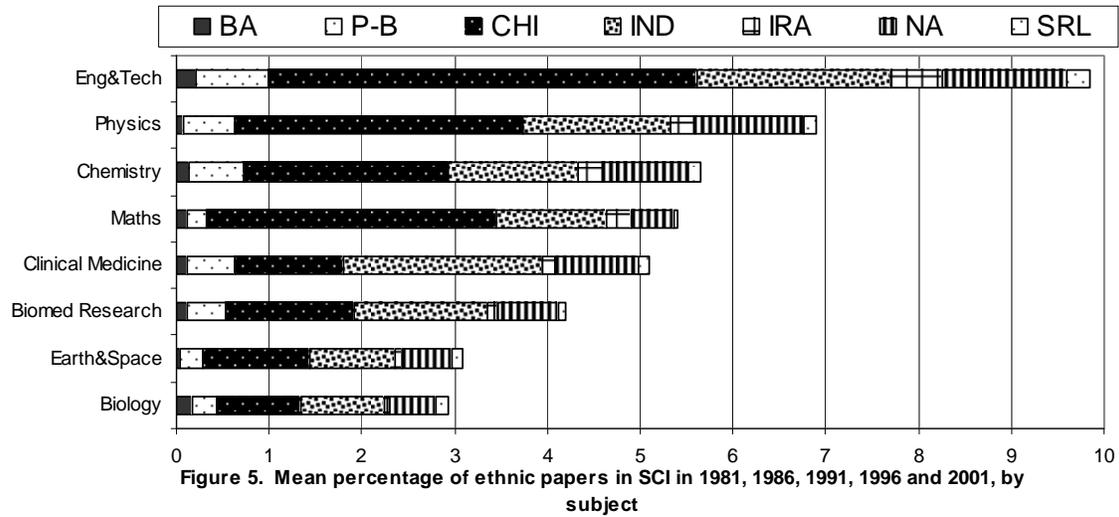
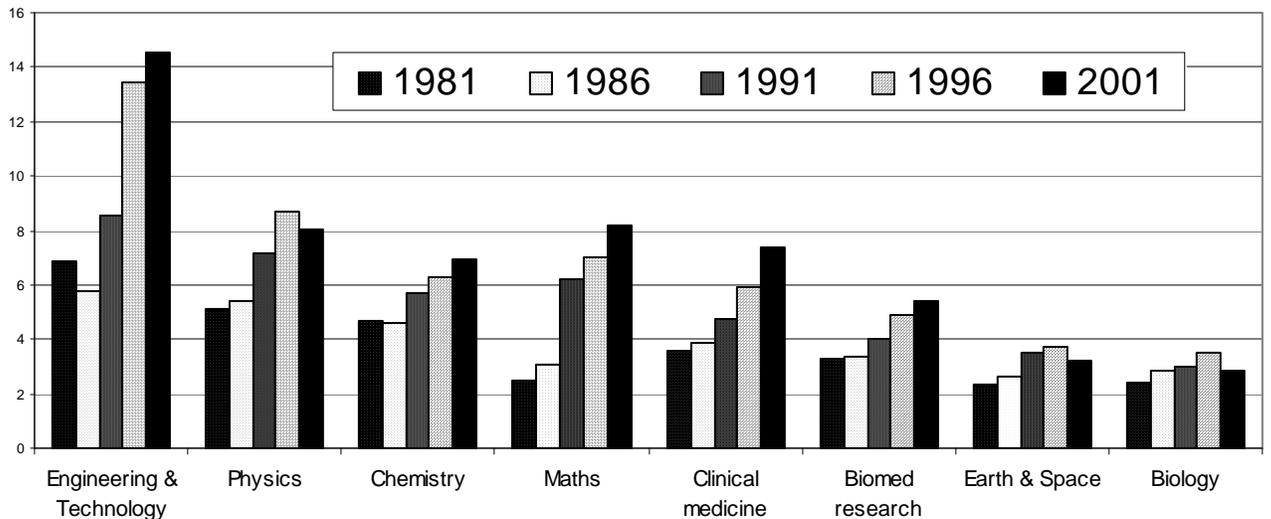


Figure 6 shows the percentages of ethnic-authored papers in eight major fields in five studied years. Mathematics, engineering and technology show the most dramatic growth in ethnic output of scholarly publications between 1981 and 2001. Ethnic outputs in mathematics more than tripled between 1981 and 2001, while outputs in engineering and technology and clinical medicine more than doubled. Earth and space and biology had the least ethnic participation and its growth over time was small. Unsurprisingly, fields with the biggest growth were those with the strongest dominance of Chinese authors.



4.3 Type of research and potential impact

This part of the analysis aims to describe ethnic papers according to two criteria. First is their research level (RL), or the type of research which they represent (from applied technology – RL1 to basic – RL4). Second is the potential impact category (PIC) of journals in which they were published (from PIC 1 – low impact to PIC 4 – high impact).

Figure 7 illustrates the division of all UK-only papers in terms of these two variables. It shows a clear correlation between the research level and potential impact: basic papers attract more citations than applied technology papers (15% of all RL4 category papers were in PIC4 category as opposed to 2% of RL1 papers). It shows that over 80% of applied papers (RL 1 and RL2 categories) are published in low impact journals (PIC1 and PIC2 categories).

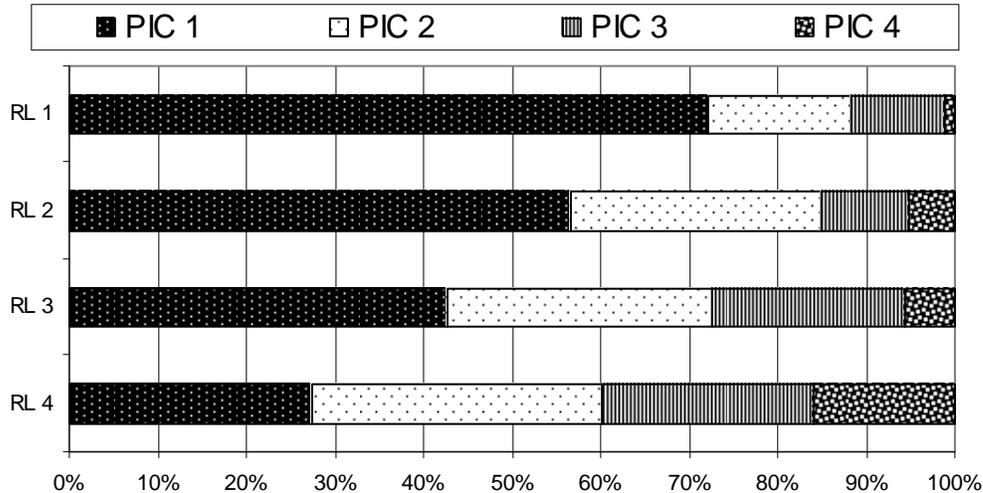


Figure 7 Percentage of UK SCI papers in 2001 in different RL and PIC categories

Figure 8 shows the differences between types of research carried out by ethnic and non-ethnic authors. While the proportion of basic research carried out by non-ethnic authors remains fairly static (around 35%), it significantly decreases for the ethnic authors (from 35% in 1981 to 23% in 2001). On the other hand, we observe a bigger increase in the proportion of RL1 category papers for ethnic minority authors (from 21% in 1981 to nearly 30% in 2001), than for papers written by non-ethnic authors (from 20% to 25%). This may be due to the fact that the numbers of ethnic-authored papers grew most dynamically in the disciplines which are primarily applied like engineering and technology (mean RL of 1.7) and clinical medicine (2.05) and less in more basic fields such as physics or chemistry (both with a mean RL value of 3.3).

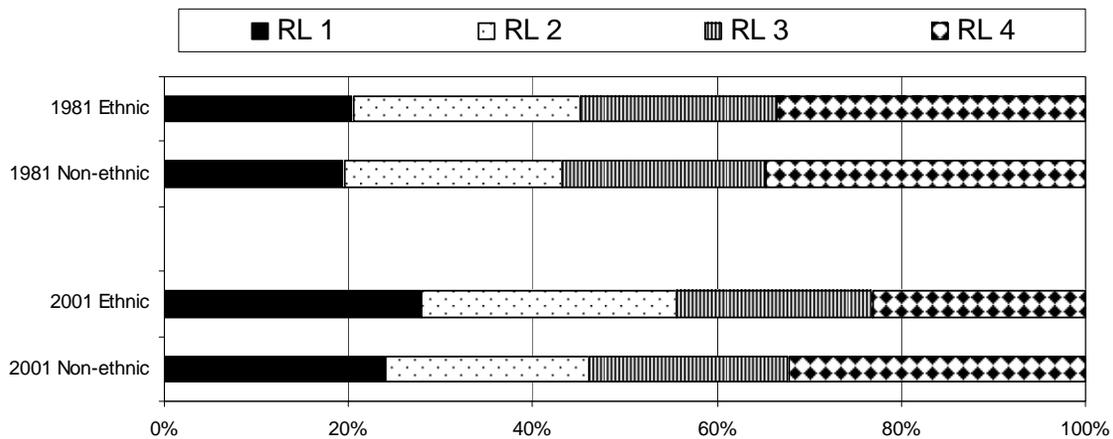


Figure 8. Ethnic and Non-ethnic UK papers by research level (RL)

Figure 9 shows the distribution of ethnic and non-ethnic papers by PIC of their journals. It shows that ethnic-authored papers achieve lower PIC scores than non-ethnic papers. Both, ethnic and non-ethnic papers show an increase in PIC values, but the rate of growth is higher for non-ethnic papers.

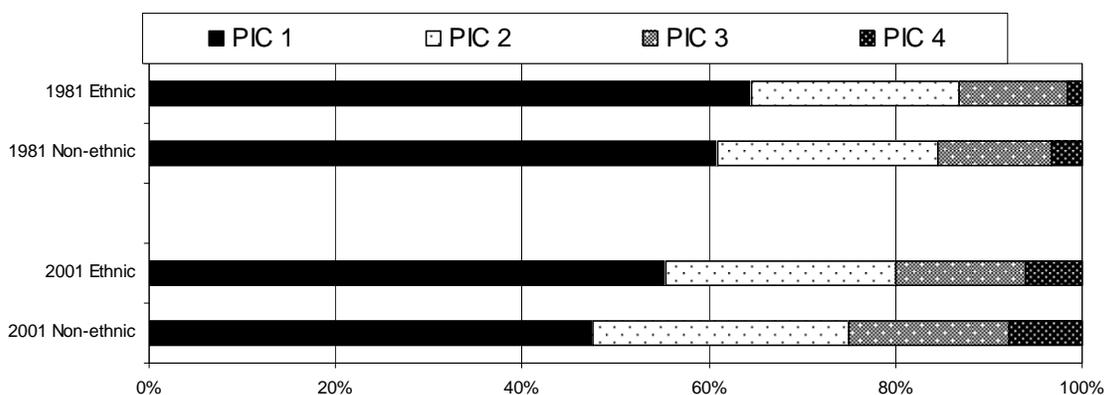


Figure 9 Ethnic and Non-ethnic UK papers by potential impact category (PIC)

It would be easy to draw a conclusion, based on the evidence in Figure 9 that ethnic-authored papers have lower potential impact than these of non-ethnic authors. However, an analysis of the characteristics of fields in which ethnic authors concentrate reveals that they are mostly applied (RL 1 and 2) and we know (see: Fig. 7) that applied papers are usually cited less frequently than basic ones, thus attracting lower PIC values.

A linear regression analysis was carried out using SPSS to determine the presence of a statistically significant relationship between a paper's PIC (dependent variable) and a series of independent variables such as number of authors (from 1 to 5 or more), number of addresses (from 1 to 5 or more), research level, field, presence of any specific ethnic group (out of seven analysed in this study) and presence of an ethnic author (irregardless of the group). The size of the sample, over 31,000 UK-only authored papers published in 2001, allows for a statistically meaningful analysis of the impact of independent variables on PIC. The results of the analysis are presented in Table 3. All variables in the left-side column have a positive impact on the PIC, while all right-hand column variables have a negative impact. The analysis showed that the presence of an ethnic author has a negative impact on the PIC (at $p > 7\%$). Amongst ethnic minority groups, only the presence of Chinese researchers had a positive impact on the PIC (despite the fact that the majority of them concentrate in low-impact fields, such as engineering and technology). Coefficients for other ethnic groups proved to be statistically not significant.

Table 3 Regression equation coefficient for potential impact category (PIC), for different input variables for UK-only SCI papers, 2001 (N = 31,969)

Variable	Coefficient	Sign.	Variable	Coefficient	Sign.
Biomedical research	0.523	0.000	Indians	-0.011	0.719
Clinical medicine	0.400	0.000	Pakistani and Bangladeshi	-0.022	0.488
Research Level	0.336	0.000	Black African	-0.054	0.355
Chinese	0.137	0.000	Ethnic	-0.084	0.007
No of authors	0.082	0.000	Earth & Space	-0.148	0.039
Sri Lankans	0.067	0.367	Chemistry	-0.232	0.001
No of addresses	0.046	0.000	Engineering & Technology	-0.342	0.000
Iranians	0.043	0.376	Physics	-0.359	0.000
North Africans	0.011	0.740	Biology	-0.397	0.000
			Mathematics	-0.629	0.000

5 Conclusions

The analysis of the presence and impact of ethnic minorities in UK science yielded some interesting results. It revealed which groups are the best represented and what is their volume of output, what type of research they favour and what is the potential impact of their publications. The analysis showed that there are big differences in the performance of various groups, with Chinese and Indians outperforming others quite significantly; that the proportion of the ethnic-authored papers grew significantly in the last twenty years; that their biggest growth was in the more applied fields of engineering and technology and clinical medicine and that they publish more, relative to non-ethnic authors, in applied and low PIC journals.

There are, however, several limitations to the surname analysis method. Firstly, we can only identify groups with distinct surnames (in the case of the UK, it was impossible to identify Afro-Caribbean authors as their surnames are frequently the same as these of English, Scottish or Welsh authors. Secondly, these groups should be relatively new to the country where marriages outside the group are not as frequent as to unduly exclude or include authors. Thirdly, the analysis excluded foreign co-authored papers (an increasingly significant part of UK published outputs) raising questions of the validity of the sample used (in the case of the UK, the percentage of foreign co-authored papers grew from 13% in 1991 to 38% in 2001).

Despite these limitations, the surname analysis showed a potential for further use. Next steps will include a sample analysis of UK foreign co-authored papers (to verify the validity of UK-only sample) and a mail survey directed at UK authors, which will be used to triangulate bibliometric data and further explain current trends.

The surname analysis method is not limited to the UK; it can be used successfully in any country that wishes to monitor the input of various ethnic groups to its scientific production.

References

- Association of University Teachers, 2000: *Ethnicity, pay and employment in higher education 1998/99*, London: Association of University Teachers.
- Carter, J./ S. Fenton/ T. Modood, 1999: *Ethnicity and employment in higher education*. London: PSI.
- Commission for Racial Equality, 1999: *Ethnic minorities in Britain*. London: Commission for Racial Equality Factsheets.
- Harding, S./ H. Dews/ S.L. Simpson, 1999: The potential to identify South Asians using a computerized algorithm to classify names, *Population Trends*, 97, pp. 46-49.
- Higher Education Funding Council for England, 2000: *Review of research*. London: HEFCE Report 00/37.
- Jones, T., 1993: *Britain's ethnic minorities*. Policy Studies Institute, London
- Lauderdale, D.S./ B. Kestenbaum, 2000: Asian American ethnic identification by surname. *Population Research and Policy Review*, v.19, pp. 283-300;
- Lewis, G., 2001: The quantity and quality of female researchers: a bibliometric study of Iceland, *Scientometrics*, v. 51(10), pp. 29-43
- McManus, I.C., 1998: Factors affecting likelihood of applicants being offered a place in medical schools in the United Kingdom in 1996 and 1997: retrospective study, *British Medical Journal*, v. 317, 24 October, pp. 1111-6.
- Mason, D., 1995: *Race and ethnicity in modern Britain*. Oxford: Oxford Univ. Press.
- Narin, F./ K.S. Hamilton, 1996: Bibliometric performance measures, *Scientometrics*, v. 36(3), pp. 293-310.
- New Scientist*, 2002: Shamefully white: it's time for science to face up to an embarrassing problem, *New Scientist*, 9 March, p. 3.
- Nicoll, A./ K. Bassett/ S.J. Ulijaszek, 1986: What's in a name: accuracy of using surnames and forenames in ascribing Asian ethnic identity in English populations' *Journal of Epidemiology and Community Health*, v. 40, pp. 364-8.
- Razum, O./ H. Zeeb/ S. Akgun, 2001: How useful is a name-based algorithm in health research among Turkish migrants in Germany? *Tropical Medicine and International Health*, v. 6(8), pp. 654-661

- Rees A./ D. Philips, 1996: Geographical spread: the national picture. Ch. 2, in: Ratcliffe, P. ed. *Ethnicity in the 1991 Census. V. 3: Social Geography and Ethnicity in Britain: Geographical Spread, Spatial Concentration and Internal Migration*. London: HMSO; pp. 23-209.
- Scott, A./ D. Pearce/ P. Goldblatt, 2001: The sizes and characteristics of the minority ethnic populations of Great Britain – latest estimates, *Population Trends*, 105, pp. 6-13.
- Sheth, T./ M. Nargundkar/ K.Chagani/ S. Anand/ C. Nair/ S. Yusuf, 1997: Classifying ethnicity utilizing the Canadian Mortality Database, *Ethnicity & Health*, v. 2(4), pp. 287-95.
- Smith, M.G., 1986: Pluralism, race and ethnicity in selected African counties, in: Rex, J. and D. Mason, eds. *Theories of race and ethnic relations*. Cambridge: Cambridge Univ. Press.
- Van Hooydonk, G., 1997: Fractional counting of multiauthored publications: consequences for the impact of authors. *JASIS*, v. 48(10), pp. 944-5.
- Webster, B.M., 2001: Polish women in science: a bibliometric analysis of Polish science and its publications, 1980-1999, *Research Evaluation*, v. 10(3), pp. 185-194.

A revised version of this contribution has been published as:

Berenika Webster (2004). Bibliometric analysis of presence and impact of ethnic minority researchers on science in the UK. *Research Evaluation* 13(1), 69-76.