

## What's in a name?

### Accuracy of using surnames and forenames in ascribing Asian\* ethnic identity in English populations

ANGUS NICOLL, KAREN BASSETT, AND STANLEY J ULIJASZEK

*From the Department of Child Health, Queen's Medical Centre, University of Nottingham, Clifton Boulevard, Nottingham NG7 2UH*

**SUMMARY** The sensitivity, specificity, and predictive value of using characteristic names to ascribe Asian ethnic identity was tested for three data sets with different Asian populations. The technique was found to have high reliability, particularly if both first and second names were used.

#### Background

Ethnic minority populations in the United Kingdom frequently show distinctly different patterns of illness from both the indigenous population and each other.<sup>1,2</sup> Through their practice of endogamy (marriage within the cultural group)<sup>3,4</sup> and slow cultural assimilation into the general population, the community from South Asia form a very distinct grouping. They show different (usually higher) rates of mortality from the indigenous population around the time of birth,<sup>5</sup> as infants,<sup>6</sup> and as adults.<sup>1</sup>

There are also subgroups within the Asian community which experience different mortality rates.<sup>2,6</sup> Because place of birth is coded in some routine statistics, the most convenient subdivision for survey work is country of birth, Pakistan, India, East Africa, Bangladesh, etc., though the major alternative, religious ethnic group (Hindu, Moslem, and Sikh) is more meaningful among Asians themselves. Both systems show differences between subgroups.<sup>1,2</sup>

There may therefore be a strong need in epidemiological enquiries concerning mortality to ascribe accurately Asian ethnic identity or nationality to individuals. Unfortunately, this is not always done where it would be relevant. For example, the 1981 census excluded a question relating to ethnicity, and such coding in routine data sets is often incomplete. Particular problems arise when we wish to examine retrospectively data sets without ethnic coding to see whether Asian identity is important. Fortunately,

Asian names are distinctive and seem to be easily distinguishable from non-Asian. Further, most of the Asian names allow subdivision into the ethnic religious subgroups. This, plus the high Asian endogamy rate,<sup>3</sup> has led some investigators to ascribe ethnic identity by inspection of names.<sup>2</sup> However, we are not aware of any studies of the sensitivity, specificity, and predictive value of this technique. The purpose of this exercise was to investigate these parameters.

#### Materials

Three data sets from areas with different populations were used.

##### CHILD HANDICAP REGISTER

The district health authority handicap register used requires that after initial referral the health visitor or school nurse who knows the child ascribe him an ethnic group. The district has a high proportion of Asian children. Records used were for children born in 1980–84 and they had the following ethnic coding:

|           |     |     |
|-----------|-----|-----|
| Asian     | 69  | 32% |
| Non-Asian | 137 | 62% |
| Uncoded   | 12  | 5%  |

The analysis was performed on the 206 coded individuals.

##### SCHOOL GROWTH SURVEY

A survey of school children in a multi-ethnic inner city area had individually coded children into ethnic groups as follows:

- (a) coding by experienced observer at the time of measurement into categories: Moslem/Sikh/West Indian/Caucasian/mixed race;

\*An ethnic Asian is defined here as a person who, irrespective of birthplace, would identify him/herself as racially and ethnically originating from peoples indigenous to India, Pakistan, Bangladesh, and Sri Lanka. This includes Asians born in East Africa, the UK, and Mauritius but excludes whites born in India (Anglo-Indians). For convenience, they will be henceforth referred to as Asians.

*What's in a name?*

(b) coding into the same categories as (a) (excluding 'mixed race') by the parents through a confidential questionnaire.

A subset of this data set with a high Asian component was used. By the observer's coding this comprised:

|              |            |     |
|--------------|------------|-----|
| Moslem       | 172        | 60% |
| Sikh         | 28         | 10% |
| Non-Asian    | 85         | 30% |
| <b>Total</b> | <b>285</b> |     |

Not all parents had completed questionnaires but for those who had the composition was:

|              |            |     |
|--------------|------------|-----|
| Moslem       | 44         | 36% |
| Sikh         | 8          | 7%  |
| Non-Asian    | 69         | 57% |
| <b>Total</b> | <b>121</b> |     |

**STILLBIRTH CERTIFICATES**

Stillbirth certificates record the mother's country of birth. A subset from two years' certificates for a regional health authority were used. The region includes the district with the handicap register but has a number of other areas with distinctly different Asian groups. The countries of birth of mothers were:

**Asian**

|            |    |
|------------|----|
| India      | 30 |
| Pakistan   | 18 |
| Bangladesh | 34 |
| Sri Lanka  | 3  |
| Mauritius  | 4  |

Table 1 *Handicap register*

| <i>Judgement</i>                            | <i>Second names only</i> |                  |              | <i>First and second names</i> |                  |              |
|---|--------------------------|------------------|--------------|-------------------------------|------------------|--------------|
|   | <i>Asian</i>             | <i>Non-Asian</i> | <i>Total</i> | <i>Asian</i>                  | <i>Non-Asian</i> | <i>Total</i> |
| <i>Reference</i>                            |                          |                  |              |                               |                  |              |
| Asian                                       | 66                       | 4                | 70           | 68                            | 0                | 68           |
| Non-Asian                                   | 3                        | 133              | 136          | 1                             | 137              | 138          |
| <b>Total</b>                                | <b>69</b>                | <b>137</b>       | <b>206</b>   | <b>69</b>                     | <b>137</b>       | <b>206</b>   |
| Sensitivity                                 |                          | 96%              |              |                               | 99%              |              |
| Specificity                                 |                          | 97%              |              |                               | 100%             |              |
| Predictive value:                           |                          |                  |              |                               |                  |              |
| Asian                                       |                          | 94%              |              |                               | 100%             |              |
| Non-Asian                                   |                          | 98%              |              |                               | 99%              |              |
| <i>Ethnic group by measurer's judgement</i> |                          |                  |              |                               |                  |              |
| Asian                                       | 194                      | 1                | 195          | 200                           | 0                | 200          |
| Non-Asian                                   | 6                        | 84               | 90           | 0                             | 85               | 85           |
| <b>Total</b>                                | <b>200</b>               | <b>85</b>        | <b>285</b>   | <b>200</b>                    | <b>85</b>        | <b>285</b>   |
| Sensitivity                                 |                          | 97.5%            |              |                               | 100%             |              |
| Specificity                                 |                          | 98.8%            |              |                               | 100%             |              |
| Predictive value:                           |                          |                  |              |                               |                  |              |
| Asian                                       |                          | 99.5%            |              |                               | 100%             |              |
| Non-Asian                                   |                          | 93.3%            |              |                               | 100%             |              |

Numbers represent absolute numbers

|                    |            |            |
|--------------------|------------|------------|
| Total              | 7          | 23%        |
| Non-Asian          | 264        |            |
| <b>Grand total</b> | <b>343</b> | <b>77%</b> |

**Methods**

For all three data sets clinical data were removed. Then second names (surnames) alone were randomly presented to observers unfamiliar with the particular sets but experienced in working with Asian names. The observers were required to judge whether individuals were Asian or non-Asian. For the school growth study they additionally had to decide whether names were Moslem or Sikh. To aid them in both tasks they employed a 'bank' of Asian names. A "don't know" judgement was allowed, and these names were presented to an Asian expert who made a forced choice into Asian/non-Asian. The whole process was then repeated, giving the observers both first (forename) and second names.

**Results**

Analysis for each set was for sensitivity, specificity, and predictive value with the following references:

*Handicap register:*

Asian/non-Asian by health visitor, school nurse judgement (table 1)

Table 2 School growth survey

| Judgement   | Second name only<br>Reference |           |           |       | First and second names<br>Reference |       |           |       |
|---|-------------------------------|-----------|-----------|-------|-------------------------------------|-------|-----------|-------|
|   | Moslem                        | Sikh      | Non-Asian | Total | Moslem                              | Sikh  | Non-Asian | Total |
| <i>(i) Ethnic group by measurer's judgement</i>       |                               |           |           |       |                                     |       |           |       |
| Moslem  | 167                           | 5         | 1         | 173   | 170                                 | 4     | 0         | 174   |
| Sikh  | 2                             | 21        | 0         | 23    | 2                                   | 24    | 0         | 26    |
| Non-Asian   | 3                             | 2         | 84        | 89    | 0                                   | 0     | 85        | 85    |
| Total   | 172                           | 28        | 85        | 285   | 172                                 | 28    | 85        | 285   |
| Sensitivity*  |                               | (Moslem)  | 97%       |       |                                     |       | 99%       |       |
|   |                               | (Sikh)    | 75%       |       |                                     |       | 86%       |       |
| Specificity†  |                               | (Moslem)  | 93%       |       |                                     |       | 96%       |       |
|   |                               | (Sikh)    | 98%       |       |                                     |       | 99%       |       |
| Predictive value‡                                     |                               | (Moslem)  | 97%       |       |                                     |       | 97%       |       |
|   |                               | (Sikh)    | 91%       |       |                                     |       | 92%       |       |
| <i>(ii) Ethnic group from parental questionnaire</i>  |                               |           |           |       |                                     |       |           |       |
| Judgement   | Asian                         | Non-Asian | Total     | Asian | Non-Asian                           | Total |           |       |
| Asian   | 51                            | 1         | 52        | 52    | 0                                   | 52    |           |       |
| Non-Asian   | 1                             | 68        | 69        | 0     | 69                                  | 69    |           |       |
| Total   | 52                            | 69        | 121       | 52    | 69                                  | 121   |           |       |
| Sensitivity   |                               |           | 98.1%     |       |                                     | 100%  |           |       |
| Specificity   |                               |           | 98.6%     |       |                                     | 100%  |           |       |
| Predictive value:                                     |                               |           |           |       |                                     |       |           |       |
| Asian   |                               |           | 98.1%     |       |                                     | 100%  |           |       |
| Non-Asian   |                               |           | 98.6%     |       |                                     | 100%  |           |       |
| <i>(iii) Ethnic group from parental questionnaire</i> |                               |           |           |       |                                     |       |           |       |
| Judgement   | Moslem                        | Sikh      | Non-Asian | Total | Moslem                              | Sikh  | Non-Asian | Total |
| Moslem  | 44                            | 0         | 1         | 45    | 40                                  | 0     | 0         | 44    |
| Sikh  | 0                             | 7         | 0         | 7     | 0                                   | 8     | 0         | 8     |
| Non-Asian   | 0                             | 1         | 68        | 69    | 0                                   | 0     | 69        | 69    |
| Total   | 44                            | 8         | 69        | 121   | 44                                  | 8     | 69        | 121   |
| Sensitivity*  |                               | (Moslem)  | 98%       |       |                                     |       |           |       |
|   |                               | (Sikh)    | 100%      |       |                                     |       |           |       |
| Specificity†  |                               | (Moslem)  | 97%       |       |                                     |       |           |       |
|   |                               | (Sikh)    | 99%       |       |                                     |       | All 100%  |       |
| Predictive value‡                                     |                               | (Moslem)  | 98%       |       |                                     |       |           |       |
|   |                               | (Sikh)    | 100%      |       |                                     |       |           |       |

\*Sensitivity defined as: (Total of correctly judged individuals in this Asian group) ÷ (Total of all this ethnic group individuals by reference)

†Specificity defined as: (Total of correctly judged non-Asians and the other Asian group) ÷ (Total of all individuals not in this Asian group by reference)

‡Predictive value defined as: (Total of Asians correctly judged to be in this group) ÷ (Total of Asians judged to be in this group)

Table 3 Stillbirth certificates

| Judgement         | Second names only<br>Reference |           |       | First and second names<br>Reference |           |       |
|-------------------|--------------------------------|-----------|-------|-------------------------------------|-----------|-------|
|                   | Asian                          | Non-Asian | Total | Asian                               | Non-Asian | Total |
| Asian             | 53                             | 20        | 73    | 74                                  | 14        | 88    |
| Non-Asian         | 26                             | 244       | 270   | 5                                   | 250       | 255   |
| Total             | 79                             | 264       | 343   | 79                                  | 264       | 343   |
| Sensitivity       |                                |           | 67.0% |                                     |           | 93.6% |
| Specificity       |                                |           | 92.4% |                                     |           | 94.7% |
| Predictive value: |                                |           |       |                                     |           |       |
| Asian             |                                |           | 72.6% |                                     |           | 84.0% |
| Non-Asian         |                                |           | 96.3% |                                     |           | 98.0% |

*What's in a name?*

*School growth survey:*

- 1 Moslem/Sikh/non-Asian by measurer's judgement
- 2 Moslem/Sikh/non-Asian by parents' description (table 2)

*Stillbirth certificates:*

Asian/non-Asian by mother's country of birth (table 3)

**Discussion**

All measures of sensitivity, specificity, and predictive value depend on the validity of the reference data. In this case that for the handicap register and growth survey is likely to be high. Those ascribing individuals to the Asian/non-Asian category intermarriage is more uncommon than the norm.<sup>3</sup> In the latter survey, the ascribing to Asian/non-Asian and Sikh/Muslim categories was done by one worker who lived in the area of the Asian community and had done six months' preparatory work getting to know its ethnic groupings. Reference judgements were made on the basis of knowing the family and/or the child individually, and our experience is that cases where Asian names were not associated with Asian ethnic group (or non-Asian grouping with Asian names) are so uncommon that they are more rather than less likely to be reported. There is more possibility of bias creeping in with the Sikh/Muslim distinction in the growth survey as the observer may have referred to names in making his judgement. However, the fact that the results were so similar in cases where the parents made the judgement themselves suggests that this is not a substantial source of bias.

Hence, that here was a high sensitivity, specificity, and predictive value for the handicap register and growth study data is encouraging in the confidence we can give to this technique for accurately ascribing the Asian/non-Asian distribution to individuals. As might be expected, there was a gain in accuracy in using both forename and family name. This also resulted in far fewer "don't knows" needing to be passed for expert judgement. Slightly less confidence can be expected in subdividing into ethnic groups in this case because of some Sikh names which are similar to or identical with names of the Caucasian population.

We found that some practical difficulties arose during this study and other work with Asian names. In particular, transcription by non-Asians often results in creative mis-spelling of Asian names, for example, Joashi as Jones. Some Asian second names are particularly difficult to ascribe as they are common to different groups: Choudhury and Patel can be both Hindu and Moslem, hence forenames become essential. Other Asian names, particularly among Hindus and Sikhs, are similar to or identical with

indigenous English names, for example, Rita (forename) Birdy (second-name) among Sikhs.

As with any true positive occurrence of low incidence Asian names can easily be missed when they are rare events among a run of indigenous names\*.

The less impressive results with the stillbirth certificates almost certainly resulted from a conflict between country of birth and ethnic groups in identifying Asian identity. Examination of those certificates falsely identified as Asian revealed that they were not observer 'blunders' but had characteristic Asian names, and the individuals had been born in countries known to have substantial Asian communities (East Africa, the West Indies, South Africa, etc). The fathers also had Asian names and usually either the same country of birth as the mother or they had been born in the Indian subcontinent. Country of birth and ethnic identity are classification systems that are obviously related, but separate analysis can produce revealing epidemiological differences.<sup>7</sup>

An incidental finding was that only one certificate had an Asian name, the mother having been born in this country, a finding that will change substantially in the immediate future as the Asian women born in Britain following immigration in the late 1960s and early 1970s come to childbearing age,<sup>8</sup> whereupon maternal country of birth will become a less reliable guide to Asian identity.

During the surveys we initially tried using names to identify other ethnic groups (eg, West Indians). However, such are the similarities between their and indigenous names that we concluded that this technique is of minimal value beyond the Asian population. Because of these two factors we urge for more routine coding of individual ethnic/religious identity (as distinct from country of birth) in data sets.

Finally, we should point out that the ethnic Asian 'mix' in Britain differs from city to city, different Asian subgroups being represented in different proportions. There is sufficient variation between the three studies to suggest that validity studies such as this should be done for different populations when the technique is used elsewhere.

We thank Dr Jean Richards and Professor Eva Alberman for allowing us access to the data sets; Shaama Sagar Malik and Mrs Kamla Joashi for expert advice; Mr Mark Parry-Billings and Ms Mary Braham for help in performing the study; and Mrs Susan Broadrick for expertly typing the manuscript.

\*Techniques for overcoming some of these problems and the name-bank mentioned are available (at cost of reproduction) on application to Dr. Nicoll at the correspondence address.

Miss Bassett and Mr Ilijaszek were supported by grants from the Special Trustees of the Nottingham Hospitals.

Correspondence should be directed to: Dr Angus Nicoll, Dept of Child Health, Queen's Medical Centre, Clifton Boulevard, Nottingham NG7 2UH

#### References

<sup>1</sup> Marmot M G, Adelstein A M, Bulusu L. Immigrant mortality in England and Wales, 1970-78. London: HMSO, 1984. (*Studies on medical and population subjects*, No 47).

Angus Nicoll, Karen Bassett, and Stanley J Ulijaszek

- <sup>2</sup> Balarajan R, Bulusu L, Adelstein A M, Shukla V. Patterns of mortality among migrants to England and Wales from the Indian sub-continent. *Br Med J* 1985; **289**: 1185-7.
- <sup>3</sup> Coleman D. Ethnic intermarriage in Great Britain. *Population Trends* 1985; **40**: 4-10.
- <sup>4</sup> Robinson V. Patterns of South Asian ethnic exogamy and endogamy in Britain. *Ethnic and Racial Studies* 1980; **3**, 427-43.
- <sup>5</sup> Terry P B, Condic R G, Settatee R. S. Analysis of ethnic differences in perinatal statistics. *Br Med J* 1980; **281**: 1307-8.
- <sup>6</sup> Macfarlane A, Mugford M. *Birth counts: statistics of pregnancy and childbirth*, London: HMSO, 1984.
- <sup>7</sup> Nicoll A. Congenital malformation amongst ethnic minorities. MSc Thesis, University of London (unpublished) 1985.
- <sup>8</sup> Immigrant Statistics Unit OPCS. Marriage and birth patterns among the New Commonwealth and Pakistani population. *Population Trends* 1978; **11**: 5-9.