Systematic identification of twins by computerised searches of NHS patient registers in the UK

David P Strachan, Andrew C Burnett

Abstract

Objectives – This study aimed to assess the feasibility, sensitivity, and specificity of a systematic search of the NHS central register for twins of the same sex.

Design, setting, and subjects – Computerised searches by alphanumeric NHS number, sex, and date of birth of 1.6 million patients registered with general practitioners in Kent, followed by validation by postal questionnaire sent to 66 pairs of children, 129 pairs of adult twins, and one set of triplets.

Main outcome measures – These were as follows: confirmed twin status, previous involvement in research, and willingness to participate in future research based on a twin register.

Results – The procedure was most efficient at identifying twins among children, in males born from 1930 onwards, and females born before 1940. Altogether 2397 male sets and 1684 female sets were identified as probable twins, triplets, or quadruplets with identical surnames, same dates of birth, and consecutive NHS numbers. Among a further 4004 pairs of adult females with the same dates of birth (1940–79), consecutive NHS numbers, but different (married) surnames, an estimated third are twins. Response to the postal questionnaire was 60% (197/327), including only one singleton. Only 10% of responders had previously participated in medical research, but 63% (65/103) of complete sets who responded expressed a willingness to participate in a twin register.

Conclusions – A systematic search of the NHS central register could identify large numbers of British twins. The procedure is efficient, reasonably sensitive, and highly specific, if supplemented by additional information from birth records for adult females born after 1939. The potential exists to create an important new resource for twin studies in Britain.

Twins offer unusual and interesting opportunities for biomedical research. In the classic (and widely used) twin study design the degree of similarity (concordance) of identical twins (presumed monozygotic) and unidentical twins of the same sex (presumed dizygotic) are compared. This approach indicates an upper limit to the relative importance of genetic factors, in the context of the prevailing lifestyle and environment, although it has long been recognised that the method cannot rigorously prove genetic causation. However, potential uses of twins studies extend well beyond the classic twin method to molecular genetics, environmental epidemiology, developmental biology, and behavioural science. In combination, such research can address the influence of both genetic and environmental factors, and their interaction.

Dizygotic twins who are concordant or highly discordant for disease or a quantitative trait may be studied for genetic markers as a special case of the sibling-pair technique. Monozygotic pairs who are discordant for a particular condition, however rare, offer powerful evidence for non-genetic determinants of disease. Experimental studies based on monozygotic twin pairs are particularly informative for outcomes where a powerful genetic-environmental interaction is suspected. Twins differ from singletons in both intrauterine environment and upbringing. Comparison of adult characteristics of twins and singletons may point to long term effects of these unusual early experiences. Certain experiences which are unique to twins, such as the occurrence of death or disease in a co-twin, may also deserve special study.

The rarity of twins in general population samples poses a problem for research studies focusing on dichotomous disease outcomes, rather than continuously measured physiological, psychological, or other variables. Twin series based on volunteers or respondents to mass media appeals may be biased in important respects, in particular towards an excess of females and identical pairs concordant for behavioural and disease characteristics. In order to overcome these problems, national twin registers have been compiled in Sweden, Norway, Finland, and Denmark. No national twin register exists in the UK, although Britain shares with Scandinavian countries the basic ingredients required to create one – that is, a near-complete population register and identity numbers allocated at birth or in other ways which are informative with regard to twin status. Over the years a number of scientific and medical research workers have expressed an interest in the possibility of setting up such a resource. A recent review of uses of Office of Population Censuses and Surveys (OPCS) records for medical research recommended that, “OPCS looks into the demand for, and
Identification of twins by computerised register searches

97

costs of, setting up internally a twin register to be used for medical research”.

We have conducted a pilot study in the Kent Family Health Services Authority (FHSA) area to assess the feasibility, sensitivity, and specificity of identifying twin pairs of the same sex systematically by a computerised search of the FHSA register, using only information available at the NHS central register (that is, date of birth, gender, surname, and NHS number).

Methods

RELEVANCE OF CONSECUTIVE NHS NUMBERS

All persons born in the UK since February 1952 have been allocated NHS numbers at birth registration. Most twin pairs have therefore received consecutive NHS numbers, but it is also possible that unrelated singletons of the same date of birth are registered sequentially and thereby receive consecutive NHS numbers. However, pairs of males and unmarried females with the same dates of birth, consecutive NHS numbers, and identical surnames are highly likely to be twins. Although (married) surnames cannot be used to confirm the twin status of adult female pairs with the same dates of birth and consecutive NHS numbers, knowledge of the NHS number identifies the birth registration entry which can be checked clerically for the surname at birth and mother’s maiden name.

Persons already alive at the creation of the NHS central register in 1952 were usually allocated NHS numbers which were the same as wartime national registration (NR) numbers. In many cases the latter were derived from the national register compiled on 29 September 1939, or from NR numbers allocated at birth registration from 30 September 1939 onwards. The national register was compiled by household, so that pairs of twins who were living together at enumeration in 1939 received consecutive identity numbers, as did most twin pairs who were born during 1939–52.

IDENTIFICATION OF TWIN PAIRS

A computerised search of the Kent FHSA register was conducted to identify individuals of the same sex and date of birth with consecutive NHS numbers, suggesting that they might be twins. The search was further restricted in men and children to pairs with identical surnames, defined hereafter as “probable” twins. This yielded 4053 pairs who were probable twins, 26 sets of probable triplets, and two sets of probable quadruplets. A stratified sample of 196 pairs of the probable twins, representing the full range of ages and both sexes, was included in the postal questionnaire survey described below.

A further 4004 pairs of adult females of the same dates of birth with consecutive NHS numbers were considered as “possible” twins. The birth records of a sample of 512 possible female pairs were checked by staff of the NHS central register and those with identical surname and mother’s maiden name were identified.

Postal Questionnaire Survey

Ethical approval for a pilot survey of twin pairs identified by the computer search was obtained from each of the local research ethics committees covering Kent. Altogether 196 pairs of probable twins resident in the Maidstone area were mailed a one page questionnaire in December 1994, and a single reminder was sent two months later. The purposes of contacting the pairs were as follows: to confirm that they were indeed twins; to ascertain zygosity (using simple questions which have been used and validated in the Scandinavian twin registries⁵); to enquire whether they had ever taken part in medical research because they are twins; and to explore whether they would consider joining a local register of twins who could be approached for medical research projects, but would not be contacted more than once each year.

Results

NHS Numbers

Among many distinct forms of alphanumeric NHS number used in England and Wales, four are potentially informative about twin status, in that they indicate the position on the birth register or in the household at enumeration on 29 September 1939. These informative forms are ABCD123, ABCD123, ABCDE123, and ABCYYD123 (where yy is year of registration from 1988 onwards). The proportion of all NHS numbers which are potentially informative varies by birth year and gender, as shown in table 1. Most people born since 1939 have informative NHS numbers, but for prewar generations the proportion is lower, due to the loss of wartime identity numbers on entry to the armed forces and postwar national service. This is a particular problem for males born before 1930.

Computer Search

The distribution of “probable” and “possible” pairs of twins and sets of triplets or quadruplets identified by the computer search is shown in table 2. The proportion of female pairs in the “probable” group is higher for pre-1940 births because the household number allocated in 1939 was taken as evidence of twin status regardless of differences in surname. Of 512 pairs of possible female pairs checked against birth records, 144 (28%) were confirmed as twins. This proportion was higher for 384 pairs born in 1939–41, 1950–51, and 1960–61.

<table>
<thead>
<tr>
<th>Birth year*</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920-29</td>
<td>19% (12 870/67 517)</td>
<td>64% (51 972/80 841)</td>
</tr>
<tr>
<td>1930-39</td>
<td>65% (53 056/81 018)</td>
<td>77% (61 466/80 071)</td>
</tr>
<tr>
<td>1940-43</td>
<td>88% (232 578/260 913)</td>
<td>88% (228 102/260 440)</td>
</tr>
<tr>
<td>1964-78</td>
<td>95% (160 360/168 769)</td>
<td>94% (153 185/163 818)</td>
</tr>
<tr>
<td>1979-93</td>
<td>99% (147 006/149 258)</td>
<td>99% (140 160/142 012)</td>
</tr>
</tbody>
</table>

*NHS numbers are unlikely to be informative for pre-1920 cohorts, since few such twins would have lived in the same household at national registration in 1939.
Table 2 Yield of computer search for twins of the same sex in Kent Family Health Services Authority register

| Birth year | Males | | | Females | | |
|---|---|---|---|---|---|
| 1920-39 | Probable* twin pairs | Triplets or quads | Probable* twin pairs | Triplets or quads | Possible* twin pairs |
| 1940-63 | 88 | 0 | 205 | 0 | 3 |
| 1964-78 | 682 | 3 | 61 | 0 | 2115 |
| 1979-93 | 735 | 3 | 485 | 3 | 1886 |
| Total | 2384 | 13† | 918 | 12† | 0 |

* See methods section for definition of “probable” and “possible” twins.
† Includes one set of quadruplets of the same sex.

(34%) than for 128 pairs born around 1970 (11%), probably because a higher proportion of the younger twin pairs had retained their maiden names and thereby were detected as “probable” pairs on the criterion of identical surnames.

QUESTIONNAIRE SURVEY

Questionnaires were sent to the parents of 66 pairs of children (under 16, born in 1979-93) and 261 adults (129 twin pairs, one set of triplets). Thirty one pairs were born before 1940, 54 pairs were born in 1940-63, and 44 twin pairs and the triplets were born in 1964-78. Replies were received from 155 adults and the parents of 42 pairs of children. This represents an overall response rate of 60% (197/327). Twelve questionnaires were returned as incorrectly addressed. All the children and all except one of the adult respondents were confirmed as twins or triplets.

By definition, all 42 responses for children related to a complete pair of twins. Among the 154 non-singleton adult respondents, there were 60 complete pairs of twins, one complete set of triplets and 31 unpaired twins (whose co-twin did not respond). On the basis of the questionnaire responses, 34 (57%) of the complete adult twin pairs were probably monozygotic. The remainder (including the triplets) appeared to be dizygotic. Among the 42 pairs of children, 14 were monozygotic, 20 dizygotic, and eight were of uncertain zygosity.

Among the non-singleton adult responders, 93 (60%) said that they would be willing to join a local twin register, 25 (16%) refused to join, and 36 (23%) were unsure. Of the 60 complete adult pairs, a willingness to join a register was expressed by both in 30 (50%), both declined in six pairs, in nine pairs both were unsure, and in 15 discordant answers were given (11 yes/unsure; four yes/no). All three triplets agreed to join a register. The proportion of presumed monozygotic pairs in whom both agreed to join a register was 50% (17/34), very similar to the corresponding proportion of dizygotic twins and triplets (52%, 14/27).

The parents of 35 (83%) of the 42 pairs of children expressed a willingness to place their child on a register. This proportion was similar for identical pairs (86%, 12/14) and for the remainder (82%, 23/28).

Only 15 (10%) responding adults and 4 (10%) pairs of children had previously participated in any form of medical or scientific research, with 10 (7%) adults and 2 (5%) pairs of children having taken part in twin research. Respondents who expressed an interest in joining a register were generally willing to participate in all forms of research (tables 3 and 4), with the anticipated exception of blood sampling or hospital tests among children (table 4). Adult respondents who were uncertain about joining a register fell into two distinct groups. About a quarter said they would be willing to participate in a wide range of studies, including blood sampling, whereas the others were unwilling to consider anything more elaborate than a postal questionnaire (table 3).

Discussion

A major problem in conducting twin research worldwide, and in Britain in particular, has been the lack of a suitable method of identifying a large unbiased sample of twins. Our pilot study shows that a computerised search of NHS numbers, dates of birth, and surnames is an efficient method of ascertaining twin pairs of the same sex across a wide age range. The method appears to be highly specific, provided that a supplementary check on birth details is conducted for females of marriageable age born after 29 September 1939. The pilot study of checks on maiden names suggests that about a third of these adult female pairs are twins, a sufficiently high yield to justify the additional clerical searches required.

We did not evaluate the sensitivity of our procedure directly, but indirect evidence suggests that it is reasonably high. The proportion of all children (born 1979-93) identified as twins, triplets, or quadruplets of the same sex was 1.26% (3656/291270), compared with the expected proportion of 1.43% based on published national rates of live twin and higher order multiple births of the same sex.26 The proportion of older subjects identified as twins was lower, because the method, as applied in Kent, only detects pairs in whom both members are currently resident in that single FHSA. However, 23 of 30 female pairs of Kent residents recruited by media appeals to a twin study of postmenopausal bone loss21 were identified by our computer search (Spector TD, personal communication). These figures and the results in table 1 suggest that the sensitivity of our method is 75–90%, depending upon the year of birth.

The response to our postal questionnaire was similar to that obtained in other surveys based on addresses derived from FHSA records.22 These addresses are known to be inaccurate in a proportion of cases.24 We might have
Identification of twins by computerised register searches

Table 3. Willingness of adult respondents* to participate in research

<table>
<thead>
<tr>
<th>Willing to take part in:</th>
<th>Willing to join register? (No (%))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes (n=93)</td>
</tr>
<tr>
<td>Postal survey</td>
<td>91 (98)</td>
</tr>
<tr>
<td>Phone interview</td>
<td>64 (69)</td>
</tr>
<tr>
<td>Measurement†</td>
<td>72 (77)</td>
</tr>
<tr>
<td>Blood sampling</td>
<td>82 (88)</td>
</tr>
<tr>
<td>Hospital visit</td>
<td>68 (73)</td>
</tr>
</tbody>
</table>

* Excluding the one respondent who was a singleton.
† Measurements such as height, weight, and blood pressure.

Table 4. Willingness of parents to involve children in research

<table>
<thead>
<tr>
<th>Willing for child to take part in:</th>
<th>Willing to join register? (No (%))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes (n=35)</td>
</tr>
<tr>
<td>Postal survey</td>
<td>35 (100)</td>
</tr>
<tr>
<td>Phone survey</td>
<td>31 (89)</td>
</tr>
<tr>
<td>Interview</td>
<td>31 (89)</td>
</tr>
<tr>
<td>Measurement†</td>
<td>33 (94)</td>
</tr>
<tr>
<td>Blood sampling</td>
<td>14 (41)</td>
</tr>
<tr>
<td>Hospital visit</td>
<td>18 (53)</td>
</tr>
</tbody>
</table>

† Measurements such as height, weight, and blood pressure.

expected that if both members of a twin pair were mailed, the probability of identifying at least one of the pair would be somewhat higher than in a general population survey. However, this did not seem to be the case in our study.

Among adult respondents, more pairs appeared to be identical (presumed monozygotic) than unidentical, whereas in a population sample of twins of the same sex, a slight majority of dizygotic pairs would be expected. This suggests that, although the ascertainment of twin pairs by computer searching is unlikely to be biased in respect of zygosity, the response of twins to a postal questionnaire may be somewhat biased towards monozygotic pairs, as tends to be the case elsewhere. In contrast, willingness to participate in a research register was similar for identical and unidentical twins.

The methods adopted for our computer search could be readily adapted to identify twins on the NHS central register (covering England and Wales) and the Scottish central register. Although the alphanumeric NHS numbers, which are informative with regard to twin status, are currently being replaced by an uninformative numeric form, the old NHS numbers will be retained at the central registers. The likely yield of a nationwide search would be approximately 250,000 pairs of the same sex surviving in 1991 (when NHS central register was computerised), of whom about 40,000 would have been born before 1940. A supplementary or confirmatory source for younger generations of twins would be computerised tapes of birth registration data on which multiple births have been flagged from 1963 onwards.

Birth registration tapes identify twins of both the same and different sexes, and the latter may be of interest for some research purposes. Although we did not evaluate the efficiency of the NHS number method for identifying twins of different sexes, we expect that it would be suitable for identifying pairs of different sexes who are children (sharing surnames) or were born in the 1930s (sharing a household in 1939). The specificity of the procedure would be poor for male-female pairs born 1940 onwards because most women would have married, and therefore a supplementary check on birth records would be required, as was the case in our study for all female pairs born 1940--79.

Full addresses are not held on either central register, but only at the FHSA in which an individual is currently registered. However, a number of research projects have contacted individuals of known NHS number via the FHSA or Scottish health board, and a similar procedure could be used to contact twins. Even if, as our pilot study indicates, there is incomplete response to such an approach, the opportunity arises to create a major new resource to facilitate future twin research in Great Britain.

We are indebted to staff of Kent FHSA and Kappe-Lambda Squared for assistance with the computer searches and preparation of the pilot mailing; and to staff of the Office of Population Censuses and Surveys, NHS Central Register, Southport, for arranging supplementary checking of maiden names.

12 Kendler KS, Holm NV. Differential enrollment in twin registries: its effect on prevalence and concordance rates.
Strachan, Burnett