METHODS OF ESTIMATING SURNAME FREQUENCY AND DISTRIBUTION

The name WHITEHOUSE, being of medium frequency, has proved useful for conducting a pilot exercise on methods of estimating surname frequency.

One aim of measuring surname frequency is to advise correspondents about the chances of a suspected relationship being correct. Put simply, as Whitehouse is a very common name in the West Midlands, the degree of proof required to make a connection in those parts is all the more exacting.

Another aim is to ascertain local frequencies and so arrive at some idea of the geographical origins of the name. A third objective is to establish a "fingerprint" of places where the families settled, as a guide to searching for wills, census returns etc. The Whitehouse fingerprint is so distinctive that it is often immediately obvious when the eye has fallen accidentally upon Whitehorn in an index.

1. The 1853 BMD method

An immediately attractive starting point was to use the GRO indexes and in particular to count births, deaths and marriages in the year 1853. This is because the Registrar General's report (dated 1856) relating to that year contains a table of the 50 commonest surnames in England and Wales combined.

A table in that report provides an estimate of the numbers of people in 1853 having these common surnames. This is divided into the total estimated population of 18,403,313 to give a frequency.   For example, there were an estimated 253,600 Smiths, which gives a frequency of 1 in 73 or 1.37%. Unfortunately, it is not stated how the Smith population was estimated.   It does not seem likely to have been by the same method by which total population was computed. That was done by using the population at the 1851 census of 17,927,609 and then adding a number representing the excess of birth registrations over death registrations.

Looking just at the total birth, marriage and death registrations in 1853, there are 18,775 Smiths, as determined by counting and confirmed by a figure elsewhere in the Registrar General's Report. Dividing the Smith registrations into the total of 1,198,008 for all names, gives a frequency of 1 in 64 which is considerably higher than the 1 in 73 based on population estimates. The same order of difference applies to other surnames. Thus, for example, Shaw, which is about the 45th commonest name in England & Wales combined, has a frequency of 1 in 504 based on the population estimate and 1 in 444 based on total birth, marriage and death registrations in 1853.

The population method is of no real use, since the methodology is unstated, while adding up birth, marriage and death registrations involves "double counting" the children born early in the year and who died later in the same year. Use of marriages is suspect, because the place of marriage may be biased in favour of brides, thus upsetting distribution information.

Basing a count on a single year is clearly not a good idea, as the Registrar General made counts for Smith and Jones from 1838 to 1854 and found big variations in the Smith to Jones ratio from year to year, varying from about 0.95:1 to 1.05:1.

Nevertheless, at least the "1853 BMD" method gives a crude idea of the commonness of a name. The Whitehouse frequency came out at 1 in 3309, which is a long way out of the top 50.
On the other hand, in the Penkridge (Staffs) registration district, the frequency was 1 in 74 (not much rarer than Smith in all England & Wales) and in Dudley it was 1 in 101, which is on a par with Williams in all England & Wales. Williams is the third commonest surname.

2. The 1861 to 1880 births method

One year of GRO indexes can have no statistical validity, but 20 years might.   Whitehouse and variants were counted in the GRO birth indexes for 1861 to 1880, totalling 5246 and the frequency calculated for each individual year by reference to total registrations in all England & Wales.   This varied between a low of 1 in 3442 and a high of 1 in 2786, with a standard deviation from the mean of 5.47%.

The West Midlands bias of the name was shown clearly in the following breakdown by registration district:

Reg. Dist.	% total Whitehouses
Dudley	22.8
West Bromwich	13.3
Wolverhampton	7.1
Aston	6.5
Birmingham	5.7
Walsall	5.2
Penkridge (=Cannock)	4.2
Stourbridge	3.2
Kings Norton	2.1
Stoke on Trent	1.2
Kidderminster	1.0
All other districts (E & W)	27.7

Further analysis by county can be misleading, because registration districts are not always what they seem. Dudley is a town in Worcestershire and therefore its registration district falls within group 6c. However, it includes the Staffordshire towns of Rowley Regis, Tipton and Sedgley and is in fact listed under Staffordshire in the Registrar General's reports, doubtless because the combined population of these places exceeds that of Dudley. This kind of confusion reaches a climax in Kings Norton (Worcs, 6c) which includes Edgbaston (Warwicks) and Harborne (Staffs).   Again, frequencies can be calculated for individual registration districts, giving 1 in 85 in Penkridge and 1 in 103 in Dudley.

Taking a 20-year period starting in 1861 enables a comparison to be made between districts in England and Scotland, remembering that Scottish registration did not start until 1855.   However, for detecting the place of origin of the name, this is an undesirably late period.   Population migration was growing apace, as can be seen by comparing the percentages of Whitehouses found in "All other districts" in the four quinquennial periods:

1861-65: 22.0%;   1866-70: 24.7%;   1871-75: 30.0%;   1876-80: 32.3%.

It is suggested that it would be better to abandon comparison with Scotland (which, anyway, has a completely different spectrum of surnames) and perform a 20-year count for, say 1841 to 1860. Waiting four years, from 1837 to 1841, allows time for the registration system to settle down.

3. The 1881 census

Whitehouse and variants gave a count of 7772 in England and Wales (from fiches). Applying the headline official figure for the population of England & Wales in the 1881 census, which is 25,974,439, gave a Whitehouse frequency of 1 in 3342, which is considerably lower than the mean of 1 in 3086 (32.50411 Whitehouses per thousand) from the 20-year GRO indexes.   There is no obvious reason for the divergence.   However, as the 1881 census is a one time-point sample, it is likely to be less accurate.

The 1881 census method really comes into its own for assessing distribution.   Populations for individual towns are available from the Official Report.   After introducing a qualification that each place must contain at least 25 Whitehouses, the following frequency table was drawn up:

Rank	Place	Frequency	Rank	Place	Frequency
1.	Cheslyn Hay	1 in 13	14.	Wednesbury	1 in 221
2.	Great Wyrley	1 in 41	15.	Stourbridge	1 in 232
3.	Tipton	1 in 48	16.	Rowley Regis	1 in 261
4.	West Bromwich	1 in 79	17.	Harborne	1 in 274
5.	Pelsall	1 in 89	18.	Walsall	1 in 285
6.	Sedgley	1 in 123	19.	Kingswinford	1 in 301
7.	Upper Swinford	1 in 123	20.	Wolverhampton	1 in 311
8.	Wednesfield	1 in 129*	21.	Handsworth	1 in 385
9.	Darlaston	1 in 129	22.	Aston	1 in 404
10.	Oldbury	1 in 131	23.	Kidderminster	1 in 553
11.	Willenhall	1 in 174	24.	Kings Norton	1 in 568
12.	Dudley	1 in 178	25.	Birmingham	1 in 592
13.	Cannock	1 in 195			

* This figure is suspect, as there are 3758 fiche entries under Wednesfield which appear to belong to Bilston and these include some Whitehouses.

The frequency for the small villages of Cheslyn Hay and Great Wyrley agree with high numbers seen in the 18th century parish registers. These two villages are in the Penkridge registration district, which had the highest frequency of Whitehouses.

Tipton, which came third, is in Dudley registration district, as are Sedgley and Rowley Regis. These four towns contained 1282 Whitehouses in a population of 142,733, a frequency of 1 in 111, which compares with 1 in 103 for Dudley registration district by the 20-year GRO index method.   The comparisons for Birmingham and Aston are 1 in 592 vs. 1 in 616 and 1 in 404 vs. 1 in 368.

End