

(advertisement)

The English Place-Name Society

For over eighty years the English Place-Name Society has been issuing its yearly volumes on the place-names of the counties of England. These publications, prepared under the General Editorship of the Honorary Director of the Survey of English Place-Names, are recognised as authoritative by scholars in many disciplines, and have proved of great value in a wide range of studies.

Research on the names of twenty-five complete counties has been published, and there are volumes for parts of Dorset, Staffordshire, Norfolk, Lincolnshire, Shropshire and Leicestershire. The fourth part of Shropshire and the third of Leicestershire have recently been published, and work on these counties, and several others, continues. It is hoped that the first volume of the County Durham survey, left nearly finished by Victor Watts when he died, will be ready for publication in 2005.

Some of the costs of research and publication are met by the subscriptions of members. An increase in membership would help to speed up the publication of further volumes. Members of the Society enjoy, in addition to a free copy of the county volume and of the Journal published during each year of their membership, the use of the Place-Names Room in the University of Nottingham, with its excellent reference library and other facilities. They may participate in the running of the Society by attendance at the Annual General Meeting, and are eligible for membership of its Council.

There is scope for further research on the place-names of all counties of England, including those already published. Proposals or enquiries, from students, academic supervisors, or private individuals, regarding individual or joint projects, will be gladly discussed by the Honorary Director of the Survey.

Details of membership, a list of the Society's publications, and further information can be obtained from:

The Secretary, English Place-Name Society, School of English Studies, University of Nottingham, Nottingham NG7 2RD.

<<http://www.nottingham.ac.uk/english/page1.htm>>

What Happened to the UK 1881 Census Surnames by 1997

Ken Tucker

Carleton University, Ottawa

1. Abstract

The paper establishes the primary reason for the apparent loss by 1997 of over two thirds of the surname types listed in the UK 1881 Census for England & Wales.

2. The Data

In my paper comparing the Forenames and Surnames of the 1881 UK Census (hereafter Census) with those of the 1998 Electoral Roll for Great Britain¹ (hereafter ER), I drew attention to the fact that of the 401,197 surnames listed in the Census only 128,970 (hereafter the *Survivors*) appeared in the ER: a shortage of 272,327 (hereafter the *Missing*).² I stated that the short fall would be the subject of another paper. This is that paper. I shall refer to the previous paper as *the previous paper*.

The GB ER comes in two forms: one for electoral purposes and another available to marketing organizations. Up to, and including, the 1998 ER these forms had the same content. Subsequent to the 1998 ER, members of the electorate, the enfranchised, have been able to opt out of the ER with no penalty. I remind readers that the field work for the ER was conducted in 1997.

The Census data, for England and Wales only, covered 26,124,585 people. I thank the UK Data Archive, and its director, Professor Kevin Schürer, for generously making the data available to me. The ER covered 47,054,569 registered voters in the GB. I thank Experian PLC for generously making the data available to me. I thank Professor Richard Webber of University College London for his facilitation with

¹ Great Britain (GB) comprises England, Scotland, and Wales. The United Kingdom (UK) comprises Great Britain and Northern Ireland.

² K. Tucker, 'The forenames and surnames from the UK 1998 Electoral Roll compared with those from the UK 1881 Census', *Nomina*, 27 (2004), 5-40.

Experian, and both him, and Professor Ed Callary of Northern Illinois University for their sampling work, and together with Dr Patrick Hanks of Berlin Brandenburg Academy of Sciences, for their informed comments along the way.

The first question to address is whether it is legitimate to compare the Census with the ER. There are a number of considerations. Since the ER data was for GB, and the Census data for England and Wales only, the catchment area for the ER is greater than that of the Census, so there is the possibility of recording a match in Scotland only. There is a possibility of a person in 1997 being under 17 years of age and having a *Missing* surname: a false *Missing* entry. The ER is not a complete set as the Census attempts to be, so there is the very real possibility that a name appearing in the Census but not in the ER could be that of the person, or persons, who simply were not registered on the ER: another false *Missing* entry. Furthermore, the Census recorded visitors to the UK, but these visitors names would not necessarily be found in the ER: again a false *Missing* entry. There are probably other reasons where they may not be compared in detail, but the sum impact of these anomalies, excepting the Scottish matching, is likely to be small when compared with the size of *Missing*: over two thirds of all the surname types in Census.

There is sharing of surname types within the UK, despite clear regional differences, that is not found, with say, a comparison of England with France. Therefore finding a match in Scotland would tell us that the matching form is a legitimate surname type and not a typographical error (hereafter *typo*). We may use the Scottish data legitimately. As we shall see, this matching process will cast larger nets later.

In light of the very large number of surname types in the *Missing* group, no attempt is made to identify for all exactly why they are missing, but only to identify the major reasons for the discrepancy and to give some indication of their size and impact.

Table 1 summarizes the numbers of surname types listed in the Census and in the ER.

Table 1: The Census and Electoral Roll Surname Types

#	Data Set-Surnames	Condition	Types
1	1881 Census	Initial State - Census	401,197
2	1881 Census	<i>Survivors</i> in 1998 ER	128,870
3	1881 Census	<i>Missing</i> : #1 - #2	272,327
4	1998 Electoral Roll	Initial State - ER	783,507
5	1998 Electoral Roll	New Surnames #4 - #2	564,637

There are of course substantial numbers of surnames in the ER that do not appear in the Census as shown in Table 1, Line 5. Some of these surnames would presumably have appeared in the complete Census data for the UK, but most are a consequence of substantial immigration to the UK in the intervening period. These new surnames are not discussed further in this paper.

3. Comparing Two Sets of Surnames

In reality what purports to be a set of surnames comprises a subset of surnames plus a non-overlapping set of typos. We may represent the Census set, and the ER set by the Venn diagrams Figure 1, and Figure 2 respectively, where the grey areas represent typos and the clear areas surnames. The areas for typos and surnames are the same for graphical convenience and do not imply that the number of typos and the number of surnames are the same. We can be reasonably confident that within the Census set and the ER set that the number of typo types is smaller than the number of surname types, and the counts represented by those types are overwhelmingly greater than the counts for the typos.

Incidentally, creating a typo does not make the owner disappear; if *Smiht* is typed rather than *Smith*, the number of people called *Smith* is the same, but the recorded count would be one less, and the miscalled one, *Smiht*, would have a count of one. The population remains the same, despite the typos and their impact on the counts. Fortunately as we shall see, the counts are relatively small.

Figure 1 Census Set

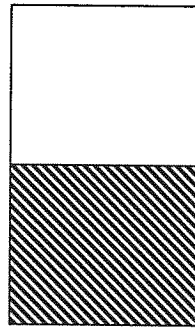
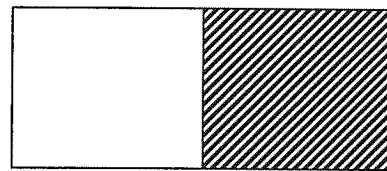


Figure 2 ER Set



When we compare Census and ER, conceptually we overlap Figures 1 and 2 to generate Figure 3.

Figure 3 Comparison of Census & ER – View A

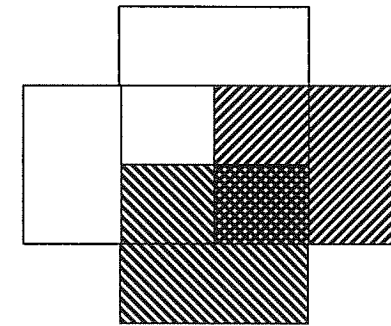
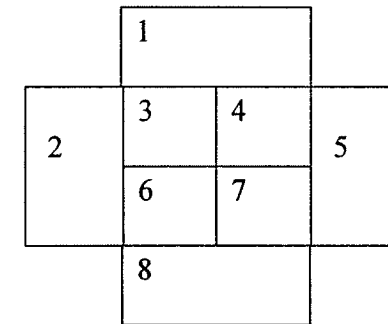


Figure 3 has eight distinct areas as shown in Figure 4.

Figure 4 Comparison of Census & ER – View B



These eight areas represent:

1. Census surnames that are unique
2. ER surnames that are unique
3. Surnames common to Census and ER
4. Census surnames matched by ER typos
5. ER typos that are unique
6. ER surnames that match Census typos
7. Census typos that match ER typos
8. Census typos that are unique

It would be ideal if we could measure these eight sectors directly but we cannot. What we can measure is the original size of Census and ER, the *Missing*, the *Survivors* and the *New* shown in Table 1.

The *Missing* are represented by areas 1 & 8

The *New* are represented by areas 2 & 5

The *Survivors* represented by areas 3, 4, 6, and 7

Typographical errors are represented in each of the groups and we are unable to detect them in a timely manner, if at all. However, we are only interested in this paper about the make up of the *Missing*; in particular, are they principally comprised of surnames, now extinct in GB, area 1, or of typos, area 8?

4. The *Missing* Surname Types

There are three major reasons why a surname appearing in the Census would not appear in the ER. The first, and most obvious, is that the surname had become extinct—no-one of that name had survived.

The second reason is that persons with that name in question elected not to register as voter, or in special cases were not old enough to be registered. The 2001 UK Census (http://www.statistics.gov.uk/census2001/population_data.asp) puts the UK population at 58,789,194 which, after subtracting the population aged from 0 to 16 inclusive, gives a prospective Electoral Roll (17 and above) of 46,161,595. This is less than the 47,054,569 entries in the ER set gathered in 1997. Assuming an increase in UK population between 1997 and 2001, it would seem that there may be duplicates in the ER set. One possible reason for this, as pointed out by Professor Webber, is that a person may have two homes and appear in the ER for both localities. Certainly, there are no grounds to suspect that the ER set is insufficiently small (as if the ER set was much smaller than the 2001 Census) to show that some surnames were definitely not included in the set. However, there is still the possibility that not all prospective voters were included and some of these possible omissions could account for some of the missing surnames. However, in my opinion the numbers of such are likely to be small in comparison to the 272,327

Missing surname types.

The third reason for a surname appearing in the Census and not in the ER would be that it had never existed in the first place. These are the typos. In large bodies of name data there are often large numbers of typos with comparatively low counts and the sum of all these errors represent a very small percentage of the population. In the case in point there are 272,327 Census surname types, 68% of the total surname types, accounting for only 3.2 percent of the Census population: an average count of less than 3. The population and counts are not inconsistent with the premise that the majority of the *Missing* are typos.

I have demonstrated in the previous paper that in the ER the fore-name *Christopher* was misspelt at least 279 times for counts generally in the ones and twos but a few of more than 100, for a total of perhaps 8,000 against the count for *Christopher* for 365,618. *Christopher* represents 1 in 280 of the forms and 98% of the count. These numbers lead one to assume that the *Missing* are probably typos and can be ignored. In order to investigate this assumption a simulator was built.

5. The Simulator

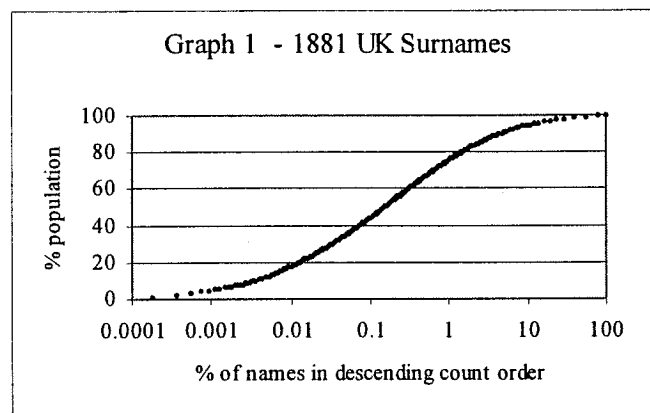
A modest typographical error rate could vastly increase the number of observed and measured types. Many typographical errors are recognizably so but others are not; a dropped letter brings *nigh nigh*. Nor is it always clear which name was the source; *nigh* could have been a corruption of *high*.

The simulation was based on an error rate of one error per period where the period was set randomly between 50 and 150 keystrokes and a range of errors: dropped letter, doubled letter, mis-keyed letter, and transposed letters, assigned randomly. The effective typographical error rate (ETER) was very close to 1%.

The goal was to take the *Survivors* set of 128,870 surname types and generate a replica of the original 401,198 UK Census database not only in terms of the number of types but also in terms of the population. If this was achieved we could say that the reduction is consistent with a modest ETER in the entire process, from the 1881 enumeration, to the creation of the electronic databases held by the UK Data Archive. Using *Survivors* as input and the ETER described

earlier, the simulation predicts that there would have been 555,242 surname types in the 1881 census results with a population of 25,302,561. This compares well with the actual case of 401,198 types and a population of 26,124,561.

Graph 1 shows cumulative population against cumulative numbers of surname types by descending count: the publishers curve, for the *Census* data. The shape is typical for surname types. If the simulated results were plotted on the same graph the reader would not see them since the second curve overlaps the first almost exactly. Hence the shape of the simulated distribution matches that of the Census. From the simulation, if all else was well, we could say that the reduction of the number of surname types is thus consistent with an ETER of less than 1%, although other factors such as name extinction are also at work.



However, matching the simulated typos with the *Missing* proved to be disappointing. The surname type matches from the Census and the Simulation numbered 146,884, but 128,870 were the *Survivors* which occur in both sets, so the real matches between the simulator, excluding the *Survivors*, and the *Missing* was only 18,014 in 272,327, about 6.7%. If typographical errors were the root cause of the *Missing* we would expect a much higher hit rate.

6. Orthography and Typos

There are probably typographical errors introduced by the Church of the Latter Day Saints' transcription process. Triple letters are virtually unknown in English and would not have been recorded and transcribed as part of the original census, yet a few appear in the transcribed file: some are shown below. They are probably mis-keys.

Abraaabrahams, 1; Bradddely, 8; Cleeeve, 1; Jffferon, 5; Mofffat, 1; Dowhagggan, 1; Haiiis, 1; McViiie, 2; Allingham, 1; Fitzsimmon, 1; Bannner, 6; Brownswood, 6; Chapppel, 1; Barrraclough, 1; Glasssodine, 1; Bilettt, 1; Blewettt, 1; Buottt, 1; Folettt, 1; Lycettt, 1; Millett, 1; Pattenden, 8; Wootton, 1; Tillettt, 1; Wittton, 1.

The number of triples per letter is: a, 1; b, 0; c, 0; d, 4; e, 5; f, 2; g, 1; h, 0; i, 2; j, 0; k, 0; l, 34; m, 2; n, 21; o, 7; p, 1; q, 1; r, 10; s, 1; t, 10; u, 0; v, 0; w, 0; x, 0; y, 0; z, 0. A total of 101 triples with over a third accounted for with one letter: the letter *Lima*.

It is strange that with only 101 'triple' surnames there are instances where the error has been made multiple time, e.g. *Jffferon, 5*. The census was enumerated person by person, but it looks as though the transcription generated the surname once for use with all the associated forenames, hence the multiple errors.

A similar list can be generated for the double vowels *ii* and *uu*, although care has to be taken as non-English names in the UK use them. There are 24 examples of the *uu* form for a total count of 51, e.g.: *Jurnez, 6*; and 67 examples of the *ii* form for a total count of 131, e.g. *Fieldiing, 5*; and *Fiielding, 2*.

It is not unlikely that the same transcription process introduced other typographical errors which are more difficult to spot, even more difficult to ascribe, and impossible to quantify.

Some of the surnames that disappeared, the *Missing*, were thus not surnames to begin with. The question is, "Are typographical errors, introduced either in the original Census or in the twentieth-century transcription, the main reason for the *Missing*?"

Identical typographical errors, which could be made in both sets of data, would allow type inflation in the Census to continue in the ER;

see Figure 4, Section 7. My own surname has been mis-spelt in correspondence as *Trucker*, by no less an authority than The British Computer Society, of which I am a long time member. I cannot find *Trucker* in any surname dictionary, but there are ten recorded in the *Census* and five in *ER*. It is easy to see why *Tucker* can be corrupted into *Trucker*, the *t* and the *r* are adjacent and both *tr* and *rt* are common combinations (two-grams) in English. I cannot rule out the possibility that *Trucker* is a surname, as it is *well-formed*. By this I mean that it obeys the rules of orthography for English surnaming, whereas, for example, *Truccker* does not. We may thus state, weakly, and unfortunately, that typos sometimes obey the rules of orthography.

Surnames from Cultural-ethnic-language groups (CELG) other than English have their own distinctive orthographies. But since the 1881 data has few new immigrant names most orthographies for the 1881 surnames are fairly well known. Thus an inspection of the names should reveal whether the surname was of the *well-formed* group, which would contain extinct surnames and well-formed typos, or of the *ill-formed* group, which would comprise all of the obvious typos. The only problem is that there are 285,000 of them which at a minute a piece would take a perfectly informed person over two years to resolve. It seemed that sampling for well-formed and ill-formed would provide a solution. Such sampling was undertaken but with mixed results. There appeared to be comparatively few obvious typos but a lot of surnames that could be typos, and the attempt to identify typographical errors was abandoned in favour of identifying other reasons for the *Missing* whilst acknowledging there were such errors. It is worth reflecting at this stage that the object is not to identify typos but to determine whether the *Missing* represent a substantial number of real people or not.

7. Root Surname Types

Closer inspection of the *Missing* showed that many were well-formed and not obvious typographical errors. Three 'roots' were arbitrarily selected, and all *Census* surname types with these roots were examined and compared with the *ER*. The roots were: *Gilder*, *Rox*, & *Tuck*. The results are presented in Tables 2, 3, and 4 respectively. I urge the

reader to cover all columns except the surname and attempt to predict which surnames are in the *Missing* group.

Table 2 *Gilder* Surname Types

SURNAME	CENSUS	MISSING	ER
Gilders	93	NO	68
Gildersberes	3	YES	
Gildersdale	1	YES	
Gildersen	9	YES	
Gildersharp	3	YES	
Gildersleaves	3	YES	
Gilderslede	1	YES	
Gildersleeve	27	NO	70
Gildersleeves	53	NO	54
Gildersleive	1	YES	
Gildersleke	2	YES	
Gildersleve	28	NO	139+1
Gildersleves	9	YES	
Gilderslewe	1	YES	
Gildersley	1	YES	
Gilderslive	5	YES	
Gildersome	1	YES	
Gilderson	38	NO	49
Gilderstein	5	YES	
Gilderstone	8	YES	
"NO" TOTAL	239	NO	280+1

Table 3 *Rox* Surname Types

SURNAME	CENSUS	MISSING	ER
Rox	17	NO	17+2
Roxall	6	NO	1
Roxbee	15	NO	26+10

SURNAME	CENSUS	MISSING	ER
Roxbernon	1	YES	
Roxberry	15	NO	13
Roxbery	4	YES	
Roxbey	1	YES	
Roxbie	1	YES	
Roxboro	2	YES	
Roxborough	46	NO	209
Roxborrow	1	YES	
Roxbourgh	7	NO	4
Roxbourough	1	YES	
Roxbrough	18	NO	40
Roxbry	1	YES	
Roxburd	2	YES	
Roxburg	17	NO	2
Roxburgh	241	NO	1387+18
Roxbury	19	NO	11
Roxby	250	NO	289+5
Roxell	1	YES	
Roxen	1	YES	
Roxes	1	YES	
Roxewall	1	YES	
Roxey	2	YES	
Roxford	2	YES	
Roxham	1	YES	
Roxhard	1	YES	
Roxher	3	YES	
Roxhole	1	YES	
Roxin	3	NO	3
Roxley	22	YES	
Roxly	4	YES	
Roxon	2	NO	5
Roxore	1	YES	

SURNAME	CENSUS	MISSING	ER
Roxson	3	YES	
Roxworthy	8	YES	
Roxy	1	NO	1
"NO" TOTAL	723	NO	1951+23

Table 4 *Tuck* Surname Types

SURNAME	CENSUS	MISSING	ER
Tuck	3829	NO	4787+25
Tuck Ingledon	3	YES	
Tuckall	1	YES	
Tuckam	2	YES	
Tuckardt	1	YES	
Tuckboll	1	YES	
Tucke	8	NO	5
Tuckee	2	NO	1
Tuckel	1	YES	
Tuckell	3	NO	9
Tucken	9	YES	
Tuckent	1	YES	
Tucker	16431	NO	21427+37
Tuckerell	1	YES	
Tuckerfield	1	YES	
Tuckeriss	3	YES	
Tuckerman	79	NO	89
Tuckermore	1	YES	
Tuckers	9	YES	
Tuckerson	1	YES	
Tuckerton	1	YES	
Tuckerworth	1	YES	
Tuckery	4	NO	1
Tuckes	3	YES	

SURNAME	CENSUS	MISSING	ER
Tucket	48	YES	
Tucketh	2	YES	
Tuckett	480	NO	504+5
Tuckey	607	NO	766+1
Tuckfield	174	NO	132
Tuckfields	4	YES	
Tuckford	1	YES	
Tuckhamel	1	YES	
Tuckhurst	1	YES	
Tuckie	1	YES	
Tuckill	5	YES	
Tucking	10	YES	
Tuckington	16	YES	
Tuckins	6	YES	
Tuckirolt	2	YES	
Tuckis	1	NO	1
Tuckiss	1	YES	
Tuckitt	5	YES	
Tuckker	3	NO	1
Tuckle	9	YES	
Tuckley	219	NO	355+2
Tuckling	2	YES	
Tuckly	2	YES	
Tuckman	28	NO	34
Tuckmann	5	YES	
Tucknell	16	NO	1
Tuckner	2	NO	4
Tuckness	2	YES	
Tuckniss	7	NO	3
Tucknoll	2	YES	
Tucknot	8	NO	1
Tucknoth	1	YES	

SURNAME	CENSUS	MISSING	ER
Tucknott	167	NO	331
Tucknutt	8	NO	60
Tuckor	7	YES	
Tuckraft	1	YES	
Tuckrell	1	YES	
Tuckridge	1	YES	
Tucks	6	YES	
Tucksbury	2	YES	
Tucksby	1	YES	
Tuckson	1	NO	1
Tucksworth	2	YES	
Tuckutt	1	YES	
Tuckwell	514	NO	836
Tuckwile	3	YES	
Tuckwill	13	YES	
Tuckwin	6	YES	
Tuckwood	192	NO	339+1
Tuckworth	5	YES	
Tuckwott	1	YES	
Tucky	8	YES	
"NO" TOTAL	23007	NO	29693+71

Uncovering the columns, we can see that each group has a high *Missing* rate of 76%, 63%, and 68% respectively. In each case the *Missing* are the low counts; this is consistent with the theory of surname survival. Of course the loss of the low counts would be consistent with them being typographical errors, but the actual surnames look like variations on a theme. Quite apart from the survival mechanics, some reduction can be assigned to the increase in literacy and automation driving out variety and increasing conformity; the adoption of the popular form.³

³ In the ER column: the "+" figures are for where the surname appears in a hyphenated name, in any position.

8. The Female Line

The Census gives the sex of those enumerated. Surnames without a male holder would be likely candidates for extinction, and all 97,530 such surnames were identified; of them, 2,415 were compound surnames and these were ignored. 82,592 of the balance of 95,130 were found in *Missing*: 86.8%. A very high and comfortable hit rate.

But what about the other 13.2%? This is one instance when the different sources is germane. In 1881 there could have been men in the UK but not in England and Wales with these surnames. Secondly, surnames with only female holders are not guaranteed extinction. It is possible for a woman with a surname Xxxxx to have a child and give the child her surname. It is also possible that a woman marry, take her husband's surname, separate, re-adopt her maiden name, produce a child and give it her surname. In such a case an extinct surname could re-appear and be perpetuated. However, it is assumed that these cases would be the exception to the rule. The net effect is that a substantial number of the *Missing*, over 30%, can be explained as being real, and now extinct, surnames.

9. Contemporary Surname Data in the USA and in Canada

We can cast our matching net wider; it is argued that if a *Missing* name can be found in other English speaking countries in reasonable numbers then it is unlikely to be a typo. No conclusion of causality is implied, or can be inferred, in this process: a *Missing* name found in quantity in the USA does not imply that the holders emigrated to the USA, but it does argue that at least the surname is real. The *Missing* were compared with 1997 lists of American Surnames and Canadian Surnames drawn from the US and Canadian Telephone directories.⁴ There are 46,356 matches with the US data, and 12,688 with the Canadian data, with a "either or both" match count of 49,155. Table 5 lists the top 100 US surnames that were matched to the *Missing*.

⁴ D. K. Tucker, 'Distribution of forenames, surnames, and forename-surname pairs in the United States', *Names*, 49 (2001), 69-96; *idem*, 'Distribution of forenames, surnames, and forename-surname pairs in Canada', *Names*, 50 (2002), 105-32.

Table 5: The Top 100 *Missing* Surnames Found in the US Telephone Directory by US Count

#	Surname	Census	US Tel
1	Braswell	5	4655
2	Cowart	3	3985
3	Grissom	2	3969
4	Suggs	20	3941
5	Shull	5	3762
6	Lofton	19	3453
7	Fugate	11	3279
8	Winstead	4	3262
9	Hulsey	5	3252
10	Weatherford	8	3199
11	Hefner	2	3086
12	Smoot	2	3068
13	Wilt	46	3027
14	Troyer	1	3006
15	Wilbanks	2	2975
16	Autry	12	2884
17	Heilman	10	2738
18	Thornburg	20	2696
19	Cupp	6	2637
20	Chaffee	3	2423
21	Musick	3	2396
22	Rife	2	2376
23	Swaim	10	2273
24	Bankston	1	2270
25	Tankersley	16	2264
26	Fritts	12	2248
27	Byrum	1	2220
28	Worsham	3	2208
29	Kinard	2	2204
30	Theriot	1	2197

#	Surname	Census	US Tel
31	Strader	1	2191
32	Ligon	3	2191
33	Koons	2	2183
34	Byerly	18	2125
35	Shupe	3	2113
36	Reavis	3	2051
37	Boring	8	2046
38	Voyles	6	1985
39	Bemis	7	1955
40	Breeding	6	1946
41	Brookshire	2	1933
42	Luster	12	1931
43	Shumway	1	1896
44	Shaner	2	1890
45	Brummett	6	1866
46	Skiles	1	1861
47	Fortney	1	1824
48	Steelman	10	1781
49	Neighbors	4	1778
50	Ketcham	1	1765
51	Gilstrap	8	1759
52	Eberly	1	1747
53	Hepler	2	1725
54	Thurber	2	1683
55	Delk	1	1674
56	Haught	5	1647
57	Thrash	16	1640
58	Wert	27	1636
59	Runnels	17	1595
60	Hollar	11	1590
61	Lingle	3	1579
62	Puryear	1	1577

#	Surname	Census	US Tel
63	Winans	11	1568
64	Fullmer	7	1563
65	Frisbie	2	1545
66	Penland	4	1539
67	Shope	1	1537
68	Gober	8	1530
69	Souder	2	1526
70	Jessee	3	1522
71	Shiver	7	1517
72	Leshner	6	1489
73	Strouse	1	1472
74	Brawner	4	1470
75	Wages	3	1441
76	Gulledge	2	1439
77	Strayer	2	1438
78	Lefler	7	1426
79	Ramsdell	11	1418
80	Hunley	23	1412
81	Karns	11	1402
82	Harville	2	1398
83	Clemmer	5	1394
84	Stutts	15	1389
85	Frierson	1	1374
86	Peppers	21	1361
87	Gaudette	1	1351
88	Hartsfield	1	1346
89	Chafin	4	1345
90	Cornelison	3	1334
91	Lippincott	15	1331
92	Brittingham	5	1327
93	Hathcock	3	1310
94	Mincey	2	1308

#	Surname	Census	US Tel
95	Shuey	1	1303
96	Stamey	1	1302
97	Shotwell	8	1297
98	Mounts	11	1297
99	Sullins	28	1291
100	Garren	1	1280

Table 6 lists the top 100 Canadian surnames that were matched to the *Missing*. Many of the Canadian names matched are of French origin, but whilst there were 12,688 matches with the Canadian data there were only 2,799 matches in addition to the US matches. The total matching by this method is 49,155 or 18% of the *Missing*.

Table 6: The Top 100 *Missing* Surnames Found in the Canadian Telephone Directory by Canadian Count

#	Surname	Census	Can Tel
1	Dumais	1	1950
2	Hachey	1	1381
3	Heroux	1	1327
4	Malenfant	3	1225
5	Brideau	1	1139
6	Robidoux	2	1065
7	Gaudette	1	1003
8	Lampron	1	919
9	Doerksen	1	874
10	Roussy	2	794
11	Lacelle	1	785
12	Anctil	1	769
13	Frigon	1	758
14	Diotte	1	741
15	Bergevin	1	644
16	Pelland	1	638

#	Surname	Census	Can Tel
17	Besner	2	585
18	Pratte	2	545
19	Plett	1	500
20	Pauze	2	488
21	Boilard	2	468
22	Myre	5	455
23	Graveline	6	432
24	Daneau	2	385
25	Banman	7	379
26	Villemaire	1	376
27	Jessome	1	347
28	Olafson	2	339
29	Labine	3	339
30	Lusignan	3	339
31	Pesant	4	333
32	Alary	4	298
33	Burelle	1	277
34	Marinier	3	269
35	Magny	6	262
36	Hinse	9	252
37	Sawler	7	250
38	Lagasse	1	246
39	Coulas	1	239
40	Pregent	1	235
41	Guilmette	5	230
42	Wigle	3	228
43	Cayen	1	222
44	Fex	2	208
45	Spnard	1	207
46	Legrow	3	206
47	Longchamps	1	205
48	Belland	7	198

#	Surname	Census	Can Tel
49	Chenel	1	195
50	Hemond	1	195
51	Lamonde	7	195
52	Guevin	1	185
53	Hoffart	3	185
54	Nearing	1	175
55	Schmaltz	1	175
56	Blakney	43	170
57	Thurber	2	169
58	Swim	12	169
59	Strome	2	168
60	Watier	6	167
61	Myette	4	166
62	Aubie	1	161
63	Reist	5	160
64	Hulan	5	158
65	Bourbonniere	1	157
66	Chartre	1	156
67	Saillant	1	155
68	Shupe	3	155
69	Arenburg	1	153
70	Beardy	2	153
71	Lefler	7	152
72	Belval	1	150
73	Tyo	5	150
74	Tufford	1	148
75	Marmen	1	147
76	Nepton	5	146
77	Remus	8	144
78	Fullum	3	140
79	Brillon	4	132
80	Shink	2	126

#	Surname	Census	Can Tel
81	Milks	7	126
82	Keffer	2	125
83	Bellaire	6	123
84	Brisbois	1	121
85	Berdan	1	119
86	Messervey	17	116
87	Genesse	1	111
88	Avoine	5	109
89	Akey	9	109
90	Manary	6	107
91	Seney	15	107
92	Evanoff	4	105
93	Scherger	8	105
94	Guillotte	3	103
95	Phalen	9	103
96	Strader	1	102
97	Wry	8	102
98	Fallu	6	101
99	Laughren	10	99
100	Sentes	7	98

So far we have identified Female only, USA, and Canadian surnames that are in *Missing*. These three groups are not mutually exclusive but in total account for 122,774, or 45.1%, of *Missing*.

10. Use of WWW.Ancestry.com

Table 7 lists the top 100 surnames from *Missing* which remain unaccounted for at this stage. It is argued that the vast majority of these are well-formed and are minor variations on well-known surnames and are unlikely to be a result of typographical errors. For example: synonyms of the first five are: *Heal, Morley, Grady, Ridley, and Griffiths*.

Table 7: The Top 100 *Missing* Surnames as Yet Unaccounted For

#	Surname	Count
1	Heall	103
2	Mauley	87
3	Gradey	86
4	Ridly	82
5	Griffithes	81
6	Grunday	80
7	Ancott	80
8	Shalford	73
9	Austey	71
10	Hawkworth	70
11	Neustead	67
12	Critchly	67
13	Motham	64
14	Coutes	64
15	Candwell	64
16	Venns	62
17	Monsley	59
18	Douson	56
19	Barmister	56
20	Silitoe	55
21	MacKerall	55
22	Laugton	54
23	Cowly	54
24	Laugham	53
25	Timbrill	51
26	Spittals	51
27	Habberly	51
28	Beauland	51
29	Strangham	50
30	Luckling	50
31	Dewry	50
32	Uttly	49

#	Surname	Count
33	Tucket	48
34	Ravenshaw	48
35	Picup	48
36	Foule	48
37	Chalfield	48
38	Baggaly	48
39	Pergrine	47
40	Tanant	46
41	Southwaite	46
42	Shetton	46
43	Cropland	46
44	Bullons	46
45	McHall	45
46	Barnsly	45
47	Sturdges	44
48	Munay	44
49	Doharty	44
50	Blockridge	44
51	Bedmead	44
52	Woollorton	43
53	Studdon	43
54	Shakspear	43
55	Halfield	43
56	Fronde	43
57	Churchell	43
58	Bourdman	43
59	Willbraham	42
60	Stuckbury	42
61	Springale	42
62	Sildon	42
63	Cawle	42
64	Borcham	42
65	Slatham	41
66	Scarsbrick	41

#	Surname	Count
67	Leery	41
68	Laddington	41
69	Haukins	41
70	Franciss	41
71	Cults	41
72	Corbishly	41
73	Battersly	41
74	Garrord	40
75	Dowar	40
76	Capping	40
77	Amsworth	40
78	Whydale	39
79	Roulands	39
80	Milcham	39
81	McLoughton	39
82	Lydenham	39
83	Holdfield	39
84	Wholstenholme	38
85	Thornily	38
86	Shenty	38
87	Luckford	38
88	Hunen	38
89	Hollese	38
90	Gouldstraw	38
91	Austiss	38
92	Anmer	38
93	Allfield	38
94	Tratman	37
95	Ladgrove	37
96	Kellott	37
97	Hutchley	37
98	Gouring	37
99	Dairs	37
100	Blencow	37

Surprisingly, all these names with one exception are to be found on the web at www.ancestry.com. They are found in either US census records, or Birth, Death, and Marriage records; many of which would be from the UK as well as the USA, but which exclude the *Census* and *ER*. The exception is *Hollese* which appears to be a Dutch place-name; although no success on *ancestry.com* or in the US telephone listings, there is a poem on the web by *Kristi Hollese*. It begins to look like we can prove that many of the *Missing* were real surnames that have now disappeared in the UK.

11. Sampling Using WWW.Ancestry.com

There are 272,327 entries in *Missing* of which all but 149, 446 have been accounted for. From this group a random sample of 508, about 0.33%, was generated and compared to Ancestry.com. Ancestry.com offers a number of sources, including births, deaths, marriages, censuses, passenger lists, newspapers, periodicals and the like. Only census data were used; the spelling had to be exact; and the proximity feature was not used.

Hit rates were graded: 1 to 10 hits were graded A; 11 to 100 were graded B, and above 100 were graded C. Each of the 508 sample surnames were entered and inspected individually. Many of these surnames looked like variations on the more common, usual surnames. Often the orthography was marginal, but few were obvious typos such as: *Jnnian*, 4; *Gvodner*, 4; *Rlanschmidt*, 3; *Foione*, 3; *Llyodd*, 2; *Lawss*, 2; *McSsary*, 1; *Lamplhier*, 1; *Jhemington*, 1; *Hardwidx*, 1; *Blackksin*, 1.

The Sample results are given in Table 8.

Table 8 Sample Results

Hit Range	Number of
100+	3
11 to 100	76
1 to 10	181
All hits	260

The three sample surnames with 100+ hits were: *Friggs*, *Beann* and *Inigley*.

With the sample size the Standard Error (SE) would be:

$$\sqrt{p[1-p]/s}$$

where p is the probability of a hit and s is the sample size. In this case p is $(260)(100)/508$, or about 51.18 and s is 508. The SE is thus $\sqrt{((51.18)(48.72)/508)} = 2.2155\%$.

We can say with 99.7% confidence that actual hit rate would be the sample hit rate plus or minus three times the SE. This is: $51.18\% \pm 6.7\%$ or between 44.5% and 57.9%. In the worst case, 44.5% of the 149,446 outstanding names are likely to be found at Ancestry.com census files.

Of course the selected sources themselves would have their own typos and typo matching would be a real possibility so we will ignore all the hits of 10 and under. P now becomes 15.6%, and the SE 1.6%. We can say with 99.7% confidence that actual hit rate would be $15.6\% \pm 4.8\%$, or between 10.8% and 20.4%. In the worst case 10.8% of the 149,446 outstanding names are likely to be found at Ancestry.com census files. This is equivalent to 16,140 surname types.

12. Summary and Conclusion

The total number of *Missing* accounted for by surnames held only by females, matches with surnames in contemporary USA and Canada, and historical records held by Ancestry.com, is thus 139,021, which is 51%. I will call these 139,021 the *Extinct*. There are more *Extinct*, than the 128,870 *Survivors*, and there are probably more surname types that can be added to *Extinct*. *Extinct* represents a significant cultural loss, 139,021 surname types in 116 years, about 100 a month during that period, and that loss should be recorded.⁵

Many of the balance of the *Missing* are the type inflators: the typos. The challenge is to identify them.

⁵ Rosencrantz and Guildenstern may be dead, but their cousins Rosencrance and Gilderstein are extinct. With apologies to William Shakespeare and Tom Stoppard.

Land Drainage Records: A Source for Name Studies in East Lincolnshire

Arthur Owen
Thimbleby

Although I shall have something to say here about the Fenland, my main concern in what follows is with the coastal Marsh district in Lindsey within the historic county of Lincolnshire.¹ This extends for some fifty miles from Barton-on-Humber in the north to Wainfleet in the south, where it merges almost imperceptibly into the Fenland. Clifford Darby describes the Lindsey Marsh as follows:

It includes two types of land at different levels ... The Middle Marsh, sometimes known as the 'Clays', forms a zone some three to six miles wide flanking the Wolds. It lies between 20 and 100 feet above sea-level, and consists of an undulating boulder clay surface varied occasionally by patches of glacial sands and gravels ... The 'Marsh' proper is a coastal belt of silt lying almost entirely below 20 feet, and, of course, draining has done much to give it its present character.²

I quoted the above when introducing a volume of selected documents relating to the medieval Marsh in 1996.³ I do so again because the physical nature of the Lindsey Marsh and the problems to which this has given rise are basic to the records to be discussed here.⁴ It should be explained at this point that the original name for the authorities who

¹ This is a revised version of a paper read at the Thirteenth Annual Conference of the Society for Name Studies in Britain and Ireland, Cambridge, 26–29 March 2004.

² H. C. Darby, *The Domesday Geography of Eastern England* (Cambridge, 1952), p. 88.

³ *The Medieval Lindsey Marsh: Select Documents*, edited by A. E. B. Owen, Lincoln Record Society, 85 (1996), p. xiii.

⁴ A. E. B. Owen, 'Coastal erosion in East Lincolnshire', *Lincolnshire Historian*, 1, ix (1952), 330–41.