NAMES, Vol. 57 No. 1, March, 2009, 52–62

Increased Competition and Reduced Popularity: US Given Name Trends of the Twentieth and Early Twenty-First Centuries

D Kenneth Tucker
Carleton University, Ottawa, Canada

In this paper, I identify changes in the naming of children in the US from 1880 to 2006. I identify the frequencies of the most popular given names in this period, and provide graphs of the male and female populations by names ranked in order of popularity, together with graphs showing the cumulative percentage populations for name rank for both the twentieth century and the first six years of the current century. It seems that parents are increasingly and deliberately avoiding selecting known popular names for their children, resulting in a decline in the absolute population share of these names. The irony of this is that there is hot competition within the top echelon, which comprises the population previously held by the single most popular given name.

KEYWORDS Forenames, Given names, Twentieth-century, Competition, Avoidance, Naming behavior

Introduction

The principal data source used in this article is the US Social Security Agency's (SSA) *Popular Baby Names* website, which is fully described in the Appendix. The data include, for each decade from 1880 to 2006, and for males and females separately, the top 1000 given names together with the numbers in the population with each name and the percentage of the birth population with each name. The data are supplemented by the total number of births during each decade. The site is principally designed to meet the interests of those who track given name popularity. However, it is an excellent source for those who seek to understand the laws that parents implicitly obey in the everyday practice of naming a baby.

Percentage population share of the ten most popular given names

Table 1 lists the top ten given names, for both males and females, for 2006 — the latest year for which such data are available. It gives both the percentage share of

 DOI 10.1179/175622709X402663

TABLE 1
THE TOP TEN GIVEN NAMES RECORDED BY THE SSA IN 2006

Rank	Male	Percentage	Cumulative percentage	Female	Percentage	Cumulative percentage
1	Jacob	1.1327	1.1327	Emily	1.0267	1.0267
2	Michael	1.0308	2.1635	Emma	0.9159	1.9426
3	Joshua	1.0148	3.1783	Madison	0.8944	2.8370
4	Ethan	0.9396	4.1179	Isabella	0.8729	3.7099
5	Matthew	0.9274	5.0453	Ava	0.8139	4.5238
6	Daniel	0.9117	5.9570	Abigail	0.7502	5.2740
7	Christopher	0.9018	6.8588	Olivia	0.7412	6.0152
8	Andrew	0.9011	7.7599	Hannah	0.6950	6.7102
9	Anthony	0.8861	8.6460	Sophia	0.6473	7.3575
10	William	0.8649	9.5109	Samantha	0.5988	7.9563

births for individual names, and the cumulative percentage share, as recorded by the SSA. The table shows that the top ten male given names account for over 9.5 per cent of the 2006 male birth population, whereas the top ten female given names account for just below 8 per cent of the female birth population. There is stiffer competition among the female names than among the male names. However, it has not always been that way.

The sample size of those with the top 1000 given names

Table 2 shows the percentage of the birth population, by decade, with the 100 and the 1000 most popular given names in the SSA data. We can see that for the female given names, the sample of those with the top 1000 names dropped below 90 per cent of the birth data in the 1960s and has continued to decline. The male given names only dropped below the 90 per cent level in the 1990s but it is now following the female trend line. It is perhaps time to increase the number of names in the sample to 10,000. However, the current data are what we have, and since the minimum population represented is over 70 per cent, the general conclusions reached using this data are likely to be valid.

Percentage population share of the most popular given names

Figure 1 plots the percentage of births given the most popular given name from 1880 to 2006, for both males (represented by squares) and females (represented by crosses). Over these 127 years the share of births of the most popular given name went from 8.2 per cent to 1.1 per cent for males, and from 7.2 per cent to 1.0 per cent for females — in both cases a reduction by a factor of over seven. The most popular given names are not as popular as they were.

In Figure 1, and in the other graphs, the line describing the data on males always starts above the line for data on females. The data for the male given names and for the female given names can be divided into five sections, as shown in Tables 3 and 4.

TABLE 2
PERCENTAGE POPULATION COVERED IN THE SSA DATA

		Top 100		Top 1000	
	Decade	Male	Female	Male	Female
1	1880s	73.6	69.3	92.2	92.5
2	1890s	70.9	66.0	91.2	92.0
3	1900s	68.1	63.8	90.2	90.8
4	1910s	68.9	63.3	91.3	90.8
5	1920s	70.1	62.4	91.1	90.9
6	1930s	71.8	62.6	93.4	91.4
7	1940s	75.3	64.8	95.4	92.3
8	1950s	75.2	63.2	95.7	92.0
9	1960s	72.2	57.2	95.3	89.3
10	1970s	67.3	50.3	91.7	82.1
11	1980s	66.2	50.2	90.0	79.5
12	1990s	57.1	43.1	85.8	74.1
13	2000s	50.3	36.3	82.3	70.3

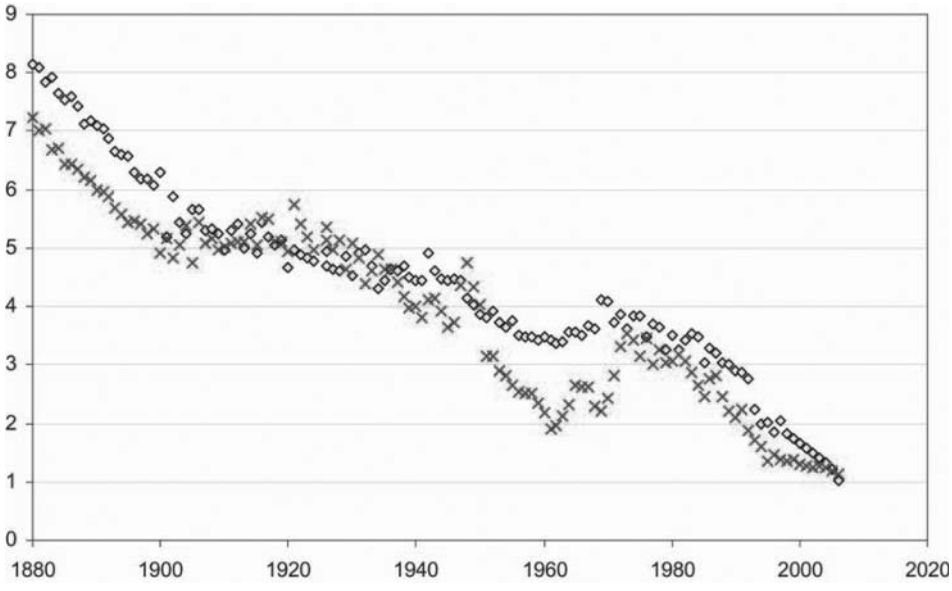

FIGURE 1 The declining popularity of the most popular given name in percentage of births (Y axis) against birth year (X axis), 1880–2006.

TABLE 3
MALE GIVEN NAMES: RISE AND FALL BY PERIOD

Period	Movement
1880–1907	Steady decline
1908–1947	Steady but reduced rate of decline
1947–1957	Resumption of steady decline
1957–1978	Gradual increase
1978–2006	Resumption of steady decline

TABLE 4
FEMALE GIVEN NAMES: RISE AND FALL BY PERIOD

Period	Movement
1880–1905	Steady decline
1906–1921	Equally steady incline
1921–1963	Resumption of steady decline
1963–1974	Precipitous decline
1974–2006	Resumption of steady decline

The five sections each cover approximately the same period for both males and females. The actual dates are a matter of interpretation; the first and the last sections are very similar for both.

The male and female graphs generally run in parallel, but the female line is the more extreme. The major differences between the two are in the second and fourth periods. In both, the female trend is up while the male trend is down. The third period for the female line was from 1921 (at 5.7 per cent) to 1963 (at 1.9 per cent), the longest decline in the period under review, albeit with an anomalous surge and decay 1946–1951. However, from 1974 the rate for both male and female sets dropped, reaching their lowest levels in 2006.

John (8.2 per cent) was the most popular male given name in 1880, but it was superseded in 2006 by *Jacob* (1.1 per cent), *Michael* (1 per cent), *Joshua* (1 per cent), *Ethan* (0.9 per cent), *Matthew* (0.9 per cent), *Christopher* (0.9 per cent), *Andrew* (0.9 per cent), and *Anthony* (0.9 per cent): nine names.

Similarly, *Mary* (7.2 per cent) was the most popular female given name in 1880, but was superseded in 2006 by *Emily* (1.0 per cent), *Emma* (0.9 per cent), *Madison* (0.9 per cent), *Isabella* (0.9 per cent), *Ava* (0.8), *Abigail* (0.8 per cent), *Olivia* (0.7 per cent), *Hannah* (0.7 per cent), *Sophia* (0.6 per cent), and *Samantha* (0.6 per cent): again, nine names.

What we see for both males and females is a striking reduction in the frequency of the most popular given name. Whereas in 1880 names like *John* and *Mary* dominated the selection, later they were displaced not by a new name but by a set of new names all of roughly the same frequency. Furthermore, the actual percentage of the population covered by the most popular given name is almost identical for both sexes.

Population against rank by sex: Zipf's law graphs

In examining the distributions of given names over the 127 years of data from the SSA, I will refer to Zipf's law (Zipf, 1949), which is found to apply to many natural events as well as to man-made systems such as naming and word use in language. Zipf's law can be framed as: X multiplied by Y equals a constant (K), which can be written as X * Y = K, where X is the rank in a descending rank series (from one, two, three, to the last rank), and Y is the number of objects at that rank.

It was stated above that *John* (at 8.2 per cent) was the most popular male given name in 1880. In this case, X = 1 (the most popular = rank one) and Y = 8.2 (the percentage of names at rank one).

If we rank the given names in descending count order and plot count against rank, each on a log scale axis, we would expect, from Zipf's law, a straight line. The diagonal line in Figure 2 represents the Zipf curve X * Y = 10,000.

The population, or count, of the top ranked name in this hypothetical case is 10,000 and the count of the hundredth ranked name is one hundred. We will take this diagonal line as a hypothetical distribution of male names. What happens if we constrain the count to a maximum of 1000 as shown by the horizontal line, which is, as we have seen, what is happening in reality? This may be thought of as 'bending' the first part of the diagonal line to the extreme, which is to the horizontal.

If the line is 'bent' we still have to find the names of the babies represented by the triangle defined by the plots (1,10,000), (1,1000), and (10,1000). We could decide to distribute them proportionally to the names ranked eleven and above, being careful not to allow any name to exceed our maximum limit of 1000. The result is that the original line shifts to the right and creates a break point at (15, 1000).

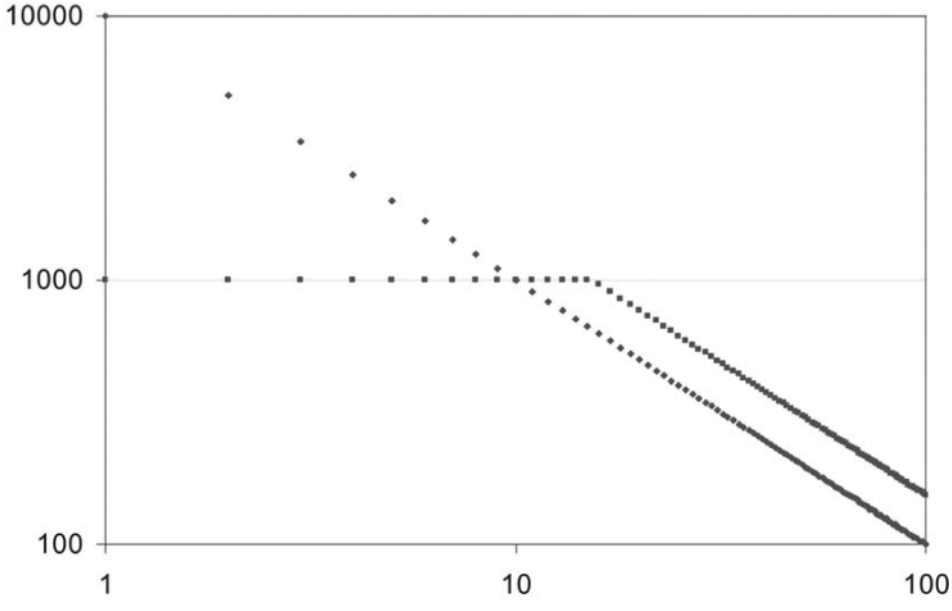

FIGURE 2 Population (Y axis) against rank (X axis) — an ideal case.

The essential fact here is that when parents discriminate against the popular given names, they force the distribution to the right of the putative Zipf line. If the amount of discrimination differs between the sexes we can expect to see the graph of the more discriminated group — in the case under discussion the females — to start at a lower rate than the males but to cross the male line and descend almost parallel with it.

The twentieth-century population by rank

Population, or count, against rank graphs can be plotted for each decade from the 1880s to the 2000s for the top 1000 male and female given names. The resultant thirteen graphs follow the same general pattern, conceptually identical to Figure 2. The male curve generally starts higher than the female curve but descends at a steeper angle and crosses the female curve; from there the two curves descend at about the same rate.

However, the graphs, like all real data, are replete with gaps, isolated plots, and the like. In the interests of clarity and economy one omnibus graph for the twentieth century, incorporating the data from the 1900s to the 1990s inclusive, for the top 1000 male and 1000 female given names, is presented in Figure 3, which is a real-life version of the idealized Figure 2. In this graph and all the subsequent graphs, the male plot line starts above the female plot line. Deaths within this set are unknown, so no claim is made that the set is truly representative of the population at the end of 1999, but it can give an overall view of practices of naming children in the twentieth century.

The three most popular given names for males were: *James* (4.8 million, 3.3 per cent), *John* (4.7 million, 3.3 per cent), and *Robert* (4.6 million, 3.2 per cent). For females they were: *Mary* (3.8 million, 2.7 per cent), *Patricia* (1.6 million, 1.1 per cent), and *Linda* (1.4 million, 1 per cent)

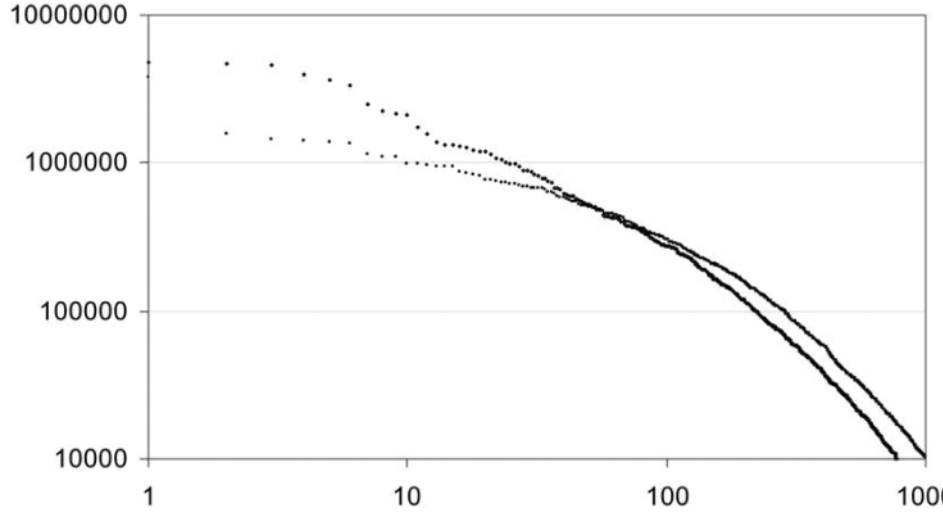

FIGURE 3 Twentieth-century population against rank.

Tucker (2004) showed that for the Great Britain 1998 electoral roll data, comparable with the data under discussion, there was a single break point at rank 104 for females and rank 59 for males. The GB plots look very similar to those considered here except that there are two break points in the US data. The graphs for the individual decades all show the same three-segment form, for both males and females. The twentieth century graph suggests a 'super-popular' group of one to five for males and one to forty-three for females, plus a popular group of six to 160 for males and forty-four to 180 for females.

We can estimate parental influence on naming practices in Figure 3 by measuring the angle between the upper part of the male and female curves and comparing it to the maximum possible — the horizontal position. In the female case the bend is about 75 per cent of the maximum possible, whereas the male bend is 47 per cent. The most that parents can do is to make the top half of the male and female lines horizontal, which has very nearly been achieved in the twenty-first century, as we shall see.

If the hypothesis that parents discriminate against popular names is true, then they did it more effectively for females than for males in the twentieth century, as the male counts are far greater than the female counts in the earlier part of the curve. For example, the sum of all the female counts up to and including rank forty-two is 38,893,064, whereas the male count is 65,821,734, so the female count at this point is only 59 per cent of the male. The total female count is 142,619,693, which is 99.7 per cent of the male count of 143,029,412, so like is being compared with like. It is thus not surprising that the male curve usually starts higher than the female curve and eventually crosses it. The only decade where the male curve does not cross the female curve is the 1900s.

The twenty-first century population by rank

Figure 4 repeats the exercise for the twenty-first century data available so far: 2001–2006. The curves for the plots for male and female given names are remarkably similar, and show severe flattening of both curves for the first fifteen or so plots.

Figure 4 contrasts sharply with the graph describing the twentieth-century data (see Figure 3). Thus the situation is perfectly described by the argument that parents are deliberately avoiding popular names. Historically, this has been more the case with females than with males, but the twenty-first century graph shows very little difference now between the sexes.

The twentieth-century cumulative population by rank

Cumulative population curves are formed by plotting for the nth name the sum of the population with names ranked from one to n. Usually, when creating cumulative curves all the names are known together with their counts, which results typically in an s-shaped curve terminating at the (100,100) point — since 100 per cent of the names must account for 100 per cent of the population.

Unfortunately, only 1000 names per decade are available from the SSA. These data will be used to approximate the top 1000 of the twentieth-century data. The first male plot happens to be *James* at 3.3 per cent; for the second plot the 3.3 per cent for *John*,

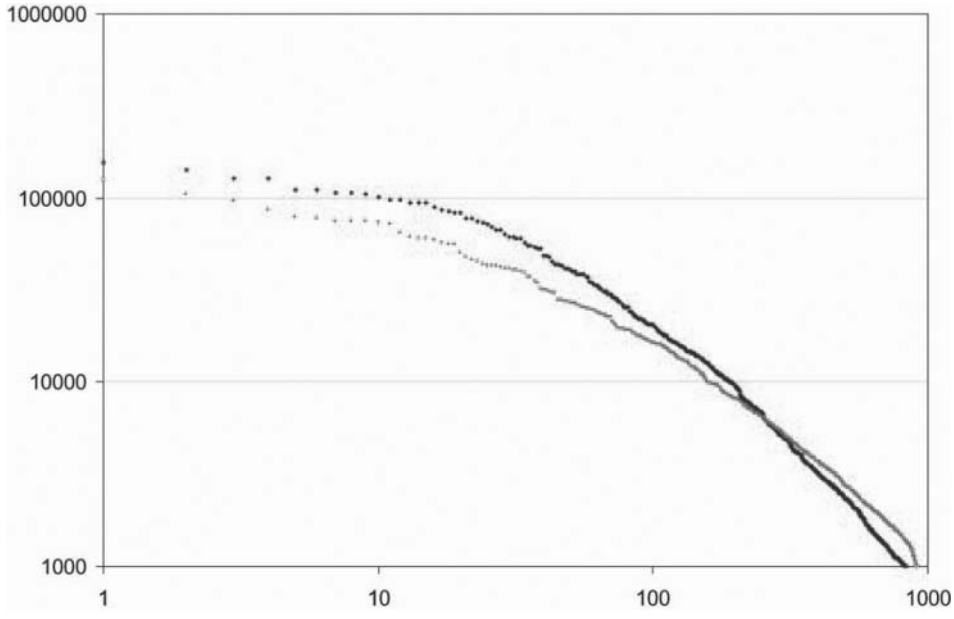

FIGURE 4 Population against rank 2001–2006.

the second name, is added to get 6.6 per cent; for the third plot 3.2 per cent is added for *Robert* to get 9.8 per cent, and so on. The female curve is created in exactly the same way. The resulting curves are shown in Figure 5.

The curves for the cumulative male population and the cumulative female population start at the same level of 3.3 per cent, but by the tenth name 25 per cent of the male population has been accounted for, whereas only 11 per cent of the female population has been covered. The female line slowly narrows the gap, and by the

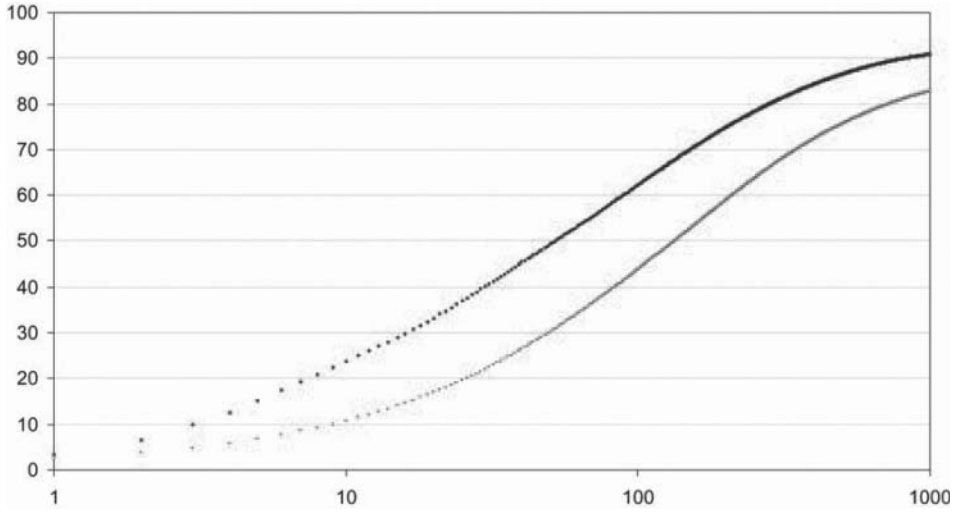

FIGURE 5 Twentieth century cumulative population against rank.

1000th name it accounts for 83 per cent of the total female population, compared with 91 per cent for the males.

The 1000th male name plotted is *Newton* with a count of 5410, whereas the 1000th female name is *Liza* with a count of 10,250. The ratio of the first to the 1000th male name is 881, but the female ratio is only 374. Clearly parents were employing one naming strategy for boys and another for girls.

The twenty-first century cumulative population by rank

It should be noted that in comparing the twentieth and twenty-first century given name data we are comparing 100 years of data with only six years. However, the first six years of the twenty-first century show that the initial bulge between the male and female curves has been removed, as shown in Figure 6.

The first to the 1000th ratios are down to 319 for the cumulative male population and 178 for the cumulative female population, showing that the differences between male and female naming practices are being reduced, with the practices for male naming moving closely to the current practices for females.

Summary

There is considerable evidence, described in this paper, that in choosing names parents in the twenty-first century are not following twentieth-century norms. Of course, this trend developed in the twentieth century, but can be seen more clearly in the twenty-first century results.

There are two basic effects. The first is that there now seems to be little difference between the process of naming boys and the process of naming girls, even ignoring

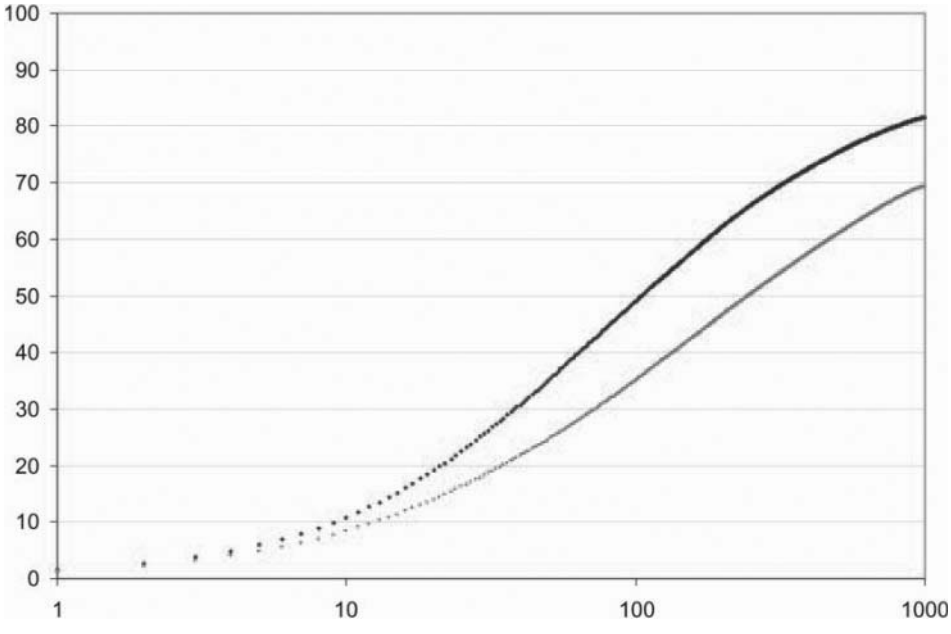

FIGURE 6 Twenty-first century cumulative population against rank, 2000–2006.

the extensive use of what were boys' names for girls. Second, there are lots of competitors for the most popular given name, resulting in a super-popular group of fewer than ten given names for both groups.

So at the beginning of the twenty-first century, the characteristics of the naming of male babies are becoming like the characteristics of the naming of female babies — popular names are becoming ever more dispersed and less concentrated.

Appendix

The principal data source for this article is the US Social Security Agency's (SSA) *Popular Baby Names* website at http://www.ssa.gov/oact/babynames

There are certain limitations on the data which the SSA wants made known to readers: see the last sentence in the SSA text below. The following is an extract from the site requested by the SSA:

> All names are from Social Security card applications for births that occurred in the United States after 1879. Names are restricted to cases where the year of birth, sex, State of birth (50 States and District of Columbia) are on record, and where the given name is at least 2 characters long. Many people born before 1937 never applied for a Social Security card, so their names are not included in our data. For others who did apply, our records may not show the place of birth, and again their names are not included in our data.
>
> All data are from a 100% sample of our records on Social Security card applications as of the end of February 2006. [As a result of early work on this article the above sentence has been changed to: '. . . end of February 2007.]
>
> Please note that name data are not edited. For example, the sex associated with a name may be incorrect. Entries such as 'Unknown' and 'Baby' are not removed from the lists.
>
> Different spellings of similar names are not combined. For example, the names Kaitlin, Kaitlyn, Kaitlynn, Katelin, Katelyn, Katelynn, and Katlyn are considered separate names and each has its own rank.
>
> When two different names are tied with the same frequency for a given year of birth, we break the tie by assigning rank in alphabetical order.
>
> Some names are applied to both males and females (for example, Shannon). Our rankings are done by sex, so that a name such as Shannon will have a different rank for males as compared to females. When you seek the popularity of a specific name (see 'Popularity of a Name'), you can also specify the sex. If you do not specify the sex, we provide rankings for the more popular name-sex combination.
>
> People using our data on popular names are urged to explicitly acknowledge the above qualifications.

On the same site can be found 'Actuarial Note #139, Name Distributions in the Social Security Area, August 1997', by Michael W. Shackleford, which lists by year the most popular male and female given names, and their percentage share of the births of that year for the appropriate gender. This list has been extended in the article using data from the main SSA site.

Bibliography

Tucker, David Kenneth. 2004. "The Forenames and Surnames from the GB 1998 Electoral Roll Compared with those from the UK 1881 Census," *Nomina* 27: 5–40.

Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

Notes on Contributor

D Kenneth Tucker is a Research Fellow at Carleton University, Ottawa, and is principally interested in the distribution of given names and surnames, and the need to embed each as objects in electronic systems to eliminate spelling errors by looking up the name rather than attempting to recreate it, which creates the real chance of skipping out of the universe of real names, as evidenced by past and contemporary censuses and the like.

Correspondence to: Dr. D. K. Tucker, 1270 Royal Palm Crescent, Manotick, Ontario K4M 1J5, Canada. Email: posthaus@igs.net