Babies' First Names: Statistical Methodology

Published on 28 March 2024

Methodology: Babies First Names

What counts as a first name for this publication?

- The statistics were based on the first name that was identified as having been recorded in the 'child's forename(s)' part of the entry of the registration of the birth. NRS identifies the names automatically, by using a computer program function which extracts (from the text in the 'child's forename(s)' field) sequences of characters which are 'delimited' by spaces (or by the start and end of the field).

- By law, all births have to be registered, and the details are sent by local registrars to the National Records of Scotland (NRS). These data allow the production of tables showing the most popular first forenames for every baby born in Scotland.

- This publication is based on births which were registered up to the point of accessing the records for analysis purposes, which is usually around early spring each year (unless their details had not been entered into the computer system by that date, which could have happened in a few cases - for example, if the registrar did not have access to the computer system, and the details were not keyed in until after this 'cut-off' date).

- These figures do not include any names that were given to babies who were stillborn.

- Our processing system will count a sequence of characters which contains a hyphen (for example 'Alice-Marie' or 'John-Paul') as a single name. However, it will count only the first word of any name that consists of two words, with a space between them. For example, the name "Da Silva" would be listed as just "Da" and St Clair would be counted as just "St". Similarly, for the purposes of these statistics, NRS would count 'J' as the first forename of a child whose forenames were recorded as 'J Arthur', and NRS would count 'JK' as the first forename if those two letters (with no intervening space) were all that was recorded in the 'child's forename(s)' field. It follows that the full lists of all the first forenames may include some entries which are not actually babies' names, and that there could be some tiny percentage errors in the analysis of the numbers of forenames given to babies. It is simply not feasible for NRS to scrutinise carefully all the babies' names that are given in a year, in order to identify those names (like 'Da Silva') that consist of two (or more) separate words, with the aim of counting them correctly for the purpose of these statistics.

- Variants based on the same name were counted separately – for example, in these statistics, 'Ben' and 'Benjamin' are different names, likewise 'Agnes' and 'Senga', and 'Tony' and 'Anthony'. Different spellings (for example Stephen, Steven; Holly, Hollie; Callum, Calum) were counted separately.

- Accents were ignored, so (for example) 'Chloe', 'Chloé', 'Chloë'. 'Chloè' and 'Chlöe' are all counted as the same name: 'Chloe'.

- In all the records that were in the NRS birth statistics database before the current statistical computer system was introduced during 2016, names were held in upper-case format (e.g. 'Mary-Frances' would be held as 'MARY-FRANCES', and both 'McKenzie' and 'Mckenzie' would be held as 'MCKENZIE'). In all the records that were added thereafter, names are held in the database in the same format as they were typed into the computer by the registrar. Therefore, to ensure that the pre- and

post-2016 data are formatted in the same way, NRS's computer programs that extract information from the database to produce statistics of baby names use a computer function to ensure that the programs process all the names in upper case format. Then, when NRS produces the tables and lists of names, it uses another computer function to convert the names that will appear in the tables and lists into so-called 'proper case' format (so that they will be more 'readable'). The method used by the latter function produces the correct result in almost all cases (for example it will convert 'MARY-FRANCES' to 'Mary-Frances'). However, in a very small percentage of cases, it cannot return a name to its exact original format. For example, all names that the programs have processed as 'MCKENZIE' will be converted to 'Mckenzie': the function will not convert some of them to 'Mckenzie' and others to 'McKenzie'. As a result, a few names in the full lists will have a lower-case letter where there should be an upper-case letter (as another example, a first forename of 'JK' would appear in those lists as 'Jk'). Please note that this issue affects only a tiny proportion of the names which appear in lists that have been produced from the statistical copy of the data, and that the administrative computer system's record of every birth registration (from which any further copies of birth certificates will be produced) has the names exactly as they were recorded by the registrar (for example, with upper-case letters only where the original entry has upper-case letters).

Information held about first names

- The amount of information about forenames that is held in the NRS birth statistics database depends upon in which of three periods the birth was registered. For births that were registered from 1974 to 1995, the 'child's forename(s)' field in the original version of that database could hold only 15 characters (including spaces between different forenames).

- Therefore, if a child was given several forenames, or some long forenames, the 'child's forename(s)' field might not have room for all of them: when that happened, the statistical database's list of that child's forenames was 'truncated' after the 15th character. In 1996, a new statistical computer system and birth statistics database were introduced, with a 'child's forename(s)' field that could hold 30 characters (including spaces) - so, for births registered from the start of 1996 to part way through 2016, a list of forenames was truncated (in the statistical database) only if there was a total of more than 30 characters (including spaces).

- The current statistical computer system and birth statistics database were introduced during 2016. Now, up to 200 characters are allowed in the 'child's forename(s)' field, so it should be able to hold all the forenames for every birth that has been registered since then. Because the previous versions of the birth statistics database could hold only a truncated list of forenames, any remaining forenames (or parts thereof) for 'pre-2016' birth registrations are unavailable for the production of statistics.

Unique names

- A first forename is counted as being 'unique' if only one birth of that sex, registered in that year, had that first forename. (Note: 'year' refers to the period up to the 'cut-off' date, in the case of the provisional figures for the latest year.) Therefore, a first forename may not be truly unique within a year. For example, a boy called Sue

might have a first forename that was unique for boys in a given year - but there could be several girls for whom Sue was their first forename. Or, a particular year might have two babies with the same 'unique' first forename: one being the only boy with that first forename, the other being the only girl. It should also be remembered that, for the purpose of these figures, a name is 'unique' as a first forename if no other birth, of the same sex, registered in the same year, has the same first forename: no account is taken of whether or not the name was given to other babies (of that sex, in that year) as, say, their second forename.

Area within Scotland

- The lists of the first forenames for each council area do not show any first forenames which were given to fewer than three babies in that area.

- When births and stillbirths are registered, the address of the usual residence of the mother at the time of the birth/stillbirth is recorded. (In a very small number of cases, this may not be known: if so, it can be counted on the basis of where the baby was found.)