Babies' First Names: points to note about the lists

This document provides some points to note about the lists, including some information about how National Records of Scotland's (NRS's) computer program identified each baby's first forename (and about how it would be unable to do so correctly in a few unusual cases).

There is a separate list for each year for which the information is available (at the time of writing, this was for each year from 1974 onwards). The list for the latest year is scheduled to be added within three months of the end of the year – for example, the list for 2017 was added in March 2018.
Please note: the publication which appears in December each year gives figures for the latest year which are based only on births which were registered in the first eleven months of that year.

The total number of babies

The total number of babies for all the names that are listed may be slightly less than the total number of births that were registered in the year concerned. The reason for this is that the lists exclude a very small percentage of cases for which the first forename was blank, due to no forenames having been recorded when a birth was registered. That may have happened for a number of reasons: for example, perhaps the parent(s) had not yet have decided what to call the baby, or perhaps the baby had died soon after birth and the parent(s) did not want to give it any forenames. The figures do not include any names that were given to babies who were stillborn.

How NRS holds the forenames

The lists for 1974 onwards were produced using the first name that was identified as having been recorded in the 'child's forenames' part of the entry for the registration of the birth in NRS's statistical database of birth records. The amount of information that is held depends upon when 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): so any first forename that was longer than that would be truncated after its 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);
- the current statistical computer system and birth statistics database, introduced during 2016, allow for up to 200 characters in the 'child's forename(s)' field.
Please note that:
- the administrative computer system's record of birth registrations was designed to hold all the names that were given, so they would all appear in full in any further copies of a child's birth certificate that may be produced;
- because 1974 is the first year for which there are records in NRS's statistical database, any lists for 1855 to 1973 that may later be produced

will have to use the first name that can be identified in the 'child's forenames' part of the entry, for the record of the registration of the birth, in the index to the Register of Births. Because the index entry records hold only the first 15 characters of the forenames, any first forename which was longer than that would be truncated to 15 characters.

How NRS identifies the first forename

Names were identified automatically by using a computer program function which extracts (from the text in the 'child's forename(s)' field part of the entry of the registration of the birth) sequences of characters which are 'delimited' by spaces (or by the start and end of the field). The computer function will count a sequence of characters which contains a hyphen (for example 'Mary-Frances') or an apostrophe (e.g. 'O'Brien') as a single name; it would count 'J' as the first forename of a child whose forenames were recorded as 'J Arthur', and 'JK' as the first forename if those two letters (with no intervening space) were all that was recorded at the start of the 'child's forename(s)' field. However, the computer program will count as two separate names any name that consists of two sequences of non-blank characters, with a space between them: for example, it will regard 'Da Silva' as two separate names ('Da' and 'Silva'), and likewise 'St Clair'. It follows that the lists may include some entries (such as 'Da' and 'St') which were only part of a baby's first forename, and that the lists will not give the full first forename in such cases. 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' and 'St Clair') that consist of two (or more) separate words, with the aim of counting them correctly for the purpose of these lists.

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'.

NRS's program that produces the lists automatically 'corrects' a few apparent 'typos' in the data that it has extracted from the database (for example, it changes any occurrences of the digits 0 and 1 to the letters O and I).

The format of the names in the lists

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 for these lists, NRS's computer programs that extract information from the database to produce lists (and 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 lists of names, it uses another computer function to convert the names that will appear in the 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, the lists will occasionally 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).

Whether NRS uses information from any re-registration of a birth

Prior to the introduction of the current statistical computer system during 2016, the NRS birth statistics database held the information that had been recorded when the birth was first registered. It did not take account of any changes that were made if a birth was re-registered (for example, to add the father's details). However, for data obtained following the introduction of the current system, in those cases where a birth was re-registered in the same calendar year as it was first registered, NRS's birth statistics database holds the information that was provided when the birth was re-registered (rather than the original registration record). This could affect the comparability of the pre- and post-2016 statistics of the numbers of forenames that were given when births were registered, if (for example) some re-registrations involve the child being given additional forenames.