

# Baby names QMI

Contact:

vsob@ons.gsi.gov.uk

Release date:

20 September 2017

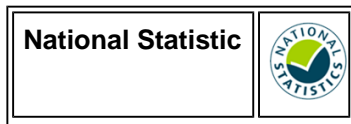
Next release:

To be announced

## Table of contents

1. [Methodology background](#)
2. [Important points about baby names data](#)
3. [Overview of the output](#)
4. [Output quality](#)
5. [About the output](#)
6. [How the output is created](#)
7. [Validation and quality assurance](#)
8. [Concepts and definitions](#)
9. [Other information](#)
10. [Sources for further information or advice](#)

# 1 . Methodology background



<b>Frequency</b>	Annual
<b>How compiled</b>	Based on third party data
<b>Geographic coverage</b>	England and Wales
<b>Last revised</b>	September 2017

## 2 . Important points about baby names data

- Baby name statistics are compiled from first names recorded when live births are registered in England and Wales as part of civil registration, a legal requirement.
- The statistics are based only on live births that occurred in the calendar year, as there is no public register of stillbirths.
- Babies born in England and Wales to women whose usual residence is outside England and Wales are included in the statistics for England and Wales as a whole, but excluded from any sub-division of England and Wales.
- The statistics are based on the exact spelling of the name given on the birth certificate; grouping names with similar pronunciation would change the rankings – exact names are given so users can group if they wish.

## 3 . Overview of the output

[Baby names](#) presents data on the first names of live-born babies. The statistics are published annually and represent births occurring in England and Wales. Figures are derived from names recorded when a birth is registered in England and Wales. The release provides counts and ranks for:

- names in England and Wales
- the top 100 names in England
- the top 100 names in Wales
- the top 10 names by month of birth
- the top 10 names by mother's region of usual residence

Office for National Statistics (ONS) took responsibility for producing baby name statistics for England and Wales in 2009. Prior to this, figures were produced by the General Register Office (GRO). A time series of baby names for [boys](#) and [girls](#) back to 1996 is available, providing counts and ranks of names in England and Wales as well as the top 10 names by month of birth. For years prior to 1996, the [top 100 rankings in England and Wales](#) were produced by GRO for 1904 to 1994 at 10-yearly intervals. These statistics are published on our website. We are unable to provide any more detailed data for these years, including counts.

For information on data quality, legislation and procedures relating to birth statistics, please see our [User Guide to Birth Statistics](#).

## 4 . Output quality

This report provides a range of information that describes the quality of the data and details any points that should be noted when using the output.

We have developed [Guidelines for measuring statistical quality](#); these are based upon the five European Statistical System (ESS) quality dimensions. This report addresses these quality dimensions and other important quality characteristics, which are:

- relevance
- timeliness and punctuality
- coherence and comparability
- accuracy
- output quality trade-offs
- assessment of user needs and perceptions
- accessibility and clarity

More information is provided about these quality dimensions in the following sections.

## 5 . About the output

### Relevance

(The degree to which statistical outputs meet users' needs.)

The annual [Baby names](#) publication presents statistics on the first names given to live-born babies born in England and Wales in a given calendar year. The release is published around 8 to 9 months after the end of the data year. Data tables are supported by a statistical bulletin providing commentary.

There is no publicly available register of stillbirths. For this reason, baby name statistics are based only on live births.

Baby name statistics do not include births to women usually resident in England or Wales who give birth abroad. They do include births that occurred in England or Wales to women whose usual residence is outside England and Wales. Such births are included in total figures for England and Wales, but are excluded from any sub-division of England and Wales.

An [interactive data visualisation tool](#) has been produced by a web developer external to Office for National Statistics (ONS) and enables users to visually compare changes in the popularity of different names since 1996.

The primary users of the data are parents and soon-to-be parents and the media. Baby name websites and those who manufacture and sell named items such as souvenir mugs also make use of the data.

Requests for more detailed information on historic baby names are often received. We are unable to provide any further information for the [years 1904 to 1995](#) beyond that available on our website. We took on the responsibility of producing baby names in 2009 and do not have the necessary data to be able to compile figures prior to 1996. For years prior to 1996, the top 100 rankings put together by the General Register Office (GRO) are published for all possible years (1904 to 1994 at 10-yearly intervals). This represents all the historic data available.

The baby name statistics are based on the exact spelling of the name given on the birth certificate. Some users request that similar names be grouped. We provide only statistics based on the exact spelling and do not group names because some groupings are not straightforward and are more a matter of opinion. Users can create their own groupings if they wish.

It is necessary to protect the confidentiality of uncommon baby names, to prevent the identification of individuals and the potential linkage of these data to other datasets, so all names with counts of fewer than three in England and Wales as a whole are redacted. Further information on the [ONS disclosure guidance for birth statistics](#) is available.

The “Assessment of user needs and perceptions” section provides further information about processes for finding out about uses and users, and their views on the [Baby names](#) release.

## Timeliness and punctuality

(Timeliness refers to the lapse of time between publication and the period to which the data refer. Punctuality refers to the gap between planned and actual publication dates.)

The annual release of Baby names is announced on the [GOV.UK release calendar](#) at least 4 weeks before publication.

Baby names are published around 8 to 9 months after the end of the data year following the full quality assurance of the data. This time lag is necessary to ensure that the statistics are based on the annual births dataset and are of the highest possible quality.

Prior to 2009, when GRO produced baby names statistics, figures were published several months earlier as they were based only on births registered in the first 46 weeks of the year. This could have introduced some seasonal bias. For example, Holly is a very popular girls name in December and a large proportion of girls called Holly born in the data year would consequently have been excluded from published figures.

[Baby names](#) for 2009 were delayed because of methodological changes to the way the baby names dataset is created. These changes were necessary to enable us to answer customer requests regarding the number of babies registered without a name.

For more details on related releases, the [GOV.UK release calendar](#) is available online and provides 12 months' advance notice of release dates. In the unlikely event of a change to the pre-announced release schedule, public attention will be drawn to the change and the reasons for the change will be explained fully at the same time, as set out in the [Code of Practice for Official Statistics](#).

## 6 . How the output is created

Baby names statistics are derived from final annual births registration data and represent all live births occurring in England and Wales in the specific calendar year. The statistics are based on the exact spelling of the name given on the birth certificate. The compilation of these statistics has been automated as much as possible to ensure efficiency. Consequently, minimal automated editing is conducted on the names. For more information on the edits applied, see the section on “Output quality trade-offs”.

## 7 . Validation and quality assurance

### Accuracy

(The degree of closeness between an estimate and the true value.)

[Baby names](#) are based on actual birth registrations provided by the General Register Office (GRO). These data represent the legal record, making it the best and most complete data source. As part of the birth registration process, before data are submitted through the Registration Online System (RON), the registrar asks the informant (typically one or both of the parents) to verify that all data entered are accurate. The registrar is then able to correct any errors. The name supplied will be the name on the birth certificate required in order to obtain a passport or a school place.

The births annual dataset used to produce the statistics is a static file of birth registration records available at the time the dataset is closed. Revisions to records can still be made after the dataset has been finalised but these will not be reflected in the annual dataset or in published statistics.

Between 1996 and 2000, the cut-off date for inclusion in the annual dataset was births occurring in the reference year that were registered by 11 February of the following year, this being 42 days after 31 December, the legal time limit for registering a birth. For 2001, the cut-off date was extended to 25 February 2002 to allow increased capture of births registered late. This change means that the annual statistics are prepared on as close to a true occurrences basis as possible without further delay to publication.

Since 2001 the annual dataset includes:

- births occurring in the reference year that were registered by 25 February the following year
- births occurring in the year prior to the reference year that were registered between 26 February in the reference year and 25 February the following year (that is, births in the previous year which had not been tabulated previously)

Prior to 2001 the annual dataset included:

- births occurring in the reference year that were registered by 11 February the following year
- births occurring in the year prior to the reference year that were registered between 12 February in the reference year and 11 February the following year (that is, births in the previous year which had not been tabulated previously)

Annual datasets for 1996 to 1999 were derived in a similar way, except that late registrations for births for all earlier years were included in the annual total, not just late registrations for births in the previous year.

## Coherence and comparability

(Coherence is the degree to which data that are derived from different sources or methods, but refer to the same topic, are similar. Comparability is the degree to which data can be compared over time and domain, for example, geographic level.)

We provide a time series of counts and ranks of baby names for [boys](#) and [girls](#) back to 1996. For years where we are unable to provide detailed data (prior to 1996), the top 100 rankings put together by GRO are published for all possible years (1904 to 1994 at 10-yearly intervals). Counts are not available before 1996, which affects the comparability of baby names prior to this date.

The published counts are based on the exact spelling of the first name given on the birth certificate. This is consistent internationally with countries such as Scotland, Northern Ireland, the Netherlands, the US, Canada and the Republic of Ireland. There are, however, some differences internationally, for example, New Zealand uses date of registration rather than date of birth.

We publish Baby names around 8 to 9 months after the end of the data year. Baby names for Scotland and Northern Ireland, published by National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA) respectively, are published in the December of the data year. NISRA finalises figures, providing more data to users around August, while NRS finalises figures a year after their initial publication at the time when the statistics for the subsequent data year are released. Office for National Statistics (ONS), NISRA and NRS all produce baby name statistics using information collected at birth registration for live births only.

We are not the only organisation to produce annual baby name statistics for England and Wales. [Bounty](#) (a parenting organisation) produces statistics using voluntary responses received from new mothers. These statistics are not as complete as those produced by us, since not all women giving birth volunteer the information to Bounty.

## 8 . Concepts and definitions

(Concepts and definitions describe the legislation governing the output and a description of the classifications used in the output.)

[Baby names](#) are derived from names recorded when a birth is registered in England and Wales. Birth registration is a legal requirement under the Births and Deaths Registration Act 1836. The registration of births occurring in England and Wales is a service carried out by the Local Registration Service in partnership with the General Register Office.

## 9 . Other information

### Output quality trade-offs

(Trade-offs are the extent to which different dimensions of quality are balanced against each other.)

Baby names are derived using information recorded at birth registration in the first forename field. The compilation of these statistics has been automated as much as possible to ensure efficiency. Consequently, minimal automated editing is conducted on the names with only the following edits being applied:

- the removal of spaces and any text following a space; text following a space is considered to be a second name rather than a first name
- the removal of accents; names are analysed without those accents since they cannot be processed within the software used to automatically generate the statistics

In the majority of cases these editing rules result in an accurate set of baby name statistics being compiled. However, there are a very small number of names in the annual datasets that are recorded in such a way that automated editing is unable to fully identify the first forename, for example:

- hyphenated names where spaces have been included between the names, for example, Amelia- Lily; such names will appear in the statistics as Amelia-
- hyphenated names where the first part of the name was included in the first forename field and the second part in the second forename field rather than all of the name being included in the first forename field; for example, Amelia- is recorded in the first forename field while Lily is recorded in the second forename field; such names will also appear in the statistics as Amelia-

In recent years, the number of hyphenated names not fully deciphered by automated editing has been very low, less than 0.01%.

Manual editing would be required to fully decipher all forenames. Manual editing is not used in the compilation of Baby names because the benefits from applying manual editing would be far outweighed by the extra time and costs associated.

## Assessment of user needs and perceptions

(The processes for finding out about uses and users, and their views on the statistical products.)

A feedback survey for Baby names took place in July 2011. The [results and responses to this survey](#) were published in August 2012.

User feedback is requested at the bottom of all emails sent by customer service teams within Vital Statistics Output Branch (VSOB).

## 10 . Sources for further information or advice

### Accessibility and clarity

(Accessibility is the ease with which users are able to access the data, also reflecting the format in which the data are available and the availability of supporting information. Clarity refers to the quality and sufficiency of the release details, illustrations and accompanying advice.)

Our recommended format for accessible content is a combination of HTML web pages for narrative, charts and graphs, with data being provided in usable formats such as CSV and Excel. We also offer users the option to download the narrative in PDF format. In some instances other software may be used, or may be available on request. Available formats for content published on our website but not produced by us, or referenced on our website but stored elsewhere, may vary. For further information, please contact us via email at [vsob@ons.gsi.gov.uk](mailto:vsob@ons.gsi.gov.uk).

For information regarding conditions of access to data, please refer to the following links:

- [Terms and conditions \(for data on the website\)](#)
- [Accessibility](#)

In addition to this Quality and Methodology Information, basic quality information relevant to each release is available in the relevant [Baby names](#) statistical bulletin.

Special extracts and tabulations of [Baby names](#) for England and Wales are available to order (subject to legal frameworks, disclosure control, resources and the [ONS charging policy](#), where appropriate). Such enquiries should be made to Vital Statistics Outputs Branch by email to [vsob@ons.gsi.gov.uk](mailto:vsob@ons.gsi.gov.uk) or by telephone on +44 (0) 1329 444110. User requested data will be published onto [our website](#).

We welcome feedback on the content, format and relevance of releases. Please send feedback to [vsob@ons.gsi.gov.uk](mailto:vsob@ons.gsi.gov.uk).

Annual baby names for 1996 onwards are available in two Excel files: one for [boys](#) and one for [girls](#). Alongside the baby names tables, a statistical bulletin is published providing supporting commentary. The bulletin outlines main findings and describes recent trends.

To aid visual interpretation further there is a baby names [interactive data visualisation tool](#). The tool shows how names have changed in popularity since 1996.

To help users identify changes within the published tables, increases in rank are denoted in blue, decreases in red and a black hyphen used to denote no change in rank. New entries into the top 100, when ranks for the year are compared with ranks for another year, are denoted by an asterisk. A colon is used to denote when a name has not previously been included in the rankings, that is, it had a count of fewer than three in the comparison year.

## Useful links

Annual [Baby names](#) are available on our website. An [interactive data visualisation tool](#) enables users to visually compare changes in the popularity of different names since 1996.

Baby names for [Scotland](#) and [Northern Ireland](#) are published by the National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA) respectively.

For information on data quality, legislation and procedures relating to birth statistics, please see our [User Guide to Birth Statistics](#).