Data Description

The Human Multiple Births Database (HMBD): An international database on twin and other multiple births

Catalina Torres
Arianna Caporali
Gilles Pison

© 2023 Catalina Torres, Arianna Caporali & Gilles Pison.

This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit. See https://creativecommons.org/licenses/by/3.0/de/legalcode.
# Contents

1  Introduction ........................................... 90

2  Data and methods ...................................... 92
   2.1 Data sources and coverage ...................... 92
   2.2 Data processing and availability ............... 93

3  Challenges with data on multiple births .......... 96

4  Database use: Example of analysis using HMBD data 98

5  Future prospects ....................................... 100

6  Acknowledgements ..................................... 100

References .................................................. 102

Appendix ..................................................... 105
The Human Multiple Births Database (HMBD):
An international database on twin and other multiple births

Catalina Torres¹
Arianna Caporali²
Gilles Pison³

Abstract

BACKGROUND
The frequency of twin births has increased dramatically since the 1970s in nearly all developed countries. This upsurge poses a public health challenge because multiple pregnancies are associated with higher health risks and other disadvantages for both the children and the parents. A better understanding of the variation and trends in twinning and other multiple rates is therefore urgently needed.

OBJECTIVE
The Human Multiple Births Database (HMBD) provides open access national statistics on multiple births for numerous countries.

METHODS
HMBD data come from the vital statistics system of each country included. We use annual counts of births by plurality to estimate the twinning and multiple birth rate for each year. All procedures performed on the input data are documented.

RESULTS
The HMBD provides the annual number of deliveries by multiplicity, the twinning rate, and the multiple rate. As of January 2023, 25 countries are included. For each country, data go back as far into the past as possible and extend until the most recent year with available data. Definitions and other specificities of each country’s data (e.g., the treatment of stillbirths in the statistics) are provided in the metadata.

¹ Eco-anthropologie (EA), Muséum national d'Histoire naturelle, CNRS, Université Paris Cité, Musée de l'Homme, Paris, France, and Institut National d’Études Démographiques, Aubervilliers, France. Email: catalina.torres@mnhn.fr.
² Institut National d'Études Démographiques, Aubervilliers, France.
³ Eco-anthropologie (EA), Muséum national d'Histoire naturelle, CNRS, Université Paris Cité, Musée de l'Homme, Paris, France, and Institut National d'Études Démographiques, Aubervilliers, France.
CONTRIBUTION
The HMBD is a unique resource, providing and documenting the most complete possible annual series of data on multiple births for each country included. All materials (data, metadata, computer codes, interactive data explorers, and supplementary material) are freely available at https://www.twinbirths.org/. At the time of writing this paper the HMBD is a work in progress, as updates and other enhancements are introduced progressively: the series for each country included is updated with data for the most recent years, and further developments in the metadata and other materials are underway.

1. Introduction

The twinning rate has increased dramatically over the last four decades in nearly all countries for which we have vital statistics information (Monden, Pison, and Smits 2021; Pison, Monden, and Smits 2015; Hoekstra et al. 2008; Pison and D’Addato 2006). For example, in the United States it has increased from 9.5 twin deliveries per 1,000 deliveries in 1975 to 16.7 in 2015 (Martin and Osterman 2019). Over the same period the twinning rate has practically doubled in many developed countries; e.g., from 9.9 to 16.1 in England and Wales, from 9.2 to 18.4 in Germany, from 9.3 to 17.5 in France, from 9.6 to 16.7 in Denmark, and from 5.9 to 9.9 in Japan. This unprecedented boom in twin births is the result of two main factors: an increase in the mean age at childbearing – as the probability of a twin pregnancy increases with the age of the mother – and a rise in the use of medically assisted reproduction.

This substantial increase in the twinning rate constitutes a public health challenge because, compared with singletons, multiple babies and their mothers are more at risk of experiencing health complications not only during pregnancy but also at delivery and in subsequent years (Johnson and Schoeni 2011; Delobel-Ayoub et al. 2009; Larroque et al. 2004). Low birth weight – associated with an increased risk of neonatal death – can result from preterm births and other complications that more commonly occur in multiple pregnancies. Stillbirth, infant, and child mortality rates are also higher for twins and other multiple children than for singletons (Monden and Smits 2017; Fresson et al. 2015; Guo and Grummer-Strawn 1993; Pison 1992). For the mothers, complications during pregnancy include increased risk of gestational diabetes mellitus, hypertensive disorders, and preeclampsia. At delivery, complications include increased risks of haemorrhage (for the mother), labour dystocia, placental abruption, umbilical cord protrusion, and asphyxia (for the baby). Furthermore, multiple-child mothers have an increased risk of post-partum depression, and parents of multiple children have an increased risk of divorce (Jena, Goldman, and Joyce 2011; Choi, Bishai, and Minkovitz 2009; Rauh-Hain et al. 2009;
Bdolah et al. 2008). A better understanding of the variation and trends in twinning and other multiple rates is therefore urgently needed.

The Human Multiple Births Database (HMBD) facilitates access to demographic data on twin and other multiple births. The database provides annual counts of deliveries by multiplicity – i.e., single and multiple deliveries – specifying, whenever possible, the number of children among the latter (twin deliveries, triplet deliveries, etc.). The twinning and multiple rates are also provided. These are defined as the annual number of twin or multiple deliveries divided by the total number of deliveries, per 1,000. The HMBD is the continuation of a database on twin births created by Gilles Pison with the collaboration of Agata D’Addato (Pison and D’Addato 2006), Christiaan Monden, and Jeroen Smits (Monden, Pison, and Smits 2021; Pison, Monden, and Smits 2015). Formerly available upon request, this dataset has been updated, extended back in time, documented, and made freely available as the HMBD at https://www.twinbirths.org. The HMBD is a work in progress, as we introduce updates in the data series by adding data for the most recent years. Moreover, further developments in the metadata and other materials are under way at the time of writing this paper, as explained further below. This paper describes the state of the HMBD as of January 2023 (HMBD - version 1). Despite being a work in progress, we decided to make the HMBD available to everyone, as it is constantly evolving and being updated as new data become available for the countries included.

Twin deliveries are the most frequent multiple deliveries; only a small fraction are triplets, quadruplets or more. For example, in France in 2010–2015, 97.9% of all multiple deliveries involved twins; the proportion of deliveries of triplets and quadruplets and more was respectively 2.0% and 0.1%. In the United States over the same period the corresponding proportions were 96.5%, 3.3%, and 0.2%. Consequently, the HMBD focuses on twin deliveries, but statistics on triplets, quadruplets, etc. are also provided where available.

In countries with reliable registration of vital events, data on twin and other multiple births are usually available through official publications of national vital statistics. Nevertheless, the readily available data on the websites of national statistics offices usually cover only the most recent years. Historical data can often be found in statistical yearbooks, bulletins, or compendiums, which may be available online as scanned copies or as hard copies in national libraries and archives. The work of searching, digitizing, compiling, documenting, and verifying long data series on the deliveries of twins and other multiple children in numerous countries and in various languages has already been carried out in the HMBD thanks to contributors and collaborators over many years. The HMBD is designed for anyone interested in the topic of twin and other multiple births. All materials related to the construction and understanding of the data are free to view,
download, and use, under the Creative Commons Licence Attribution 4.0 International (CC BY 4.0).

2. Data and methods

2.1 Data sources and coverage

The input data for the HMBD come from the vital statistics of each country included. The data cover total national populations unless otherwise indicated in the country-specific metadata files. A country’s inclusion in the HMBD depends on two main criteria. First, birth registration should be complete, which means that all or nearly all births should be declared and included in the civil registry. The completeness of birth registration is assessed based on the documentation on birth statistics provided by each country’s statistics office; occasionally we also use information found in other sources, such as research articles that discuss particular issues found in the vital registration of a given country. The second criterion is that the national statistics office should regularly publish statistics on the number of births by multiplicity, distinguishing singleton from multiple births and further specifying (if possible) multiple birth type (i.e., twins, triplets, quadruplets, etc.).

As of January 2023, the HMBD online comprises 25 national populations across Europe, America, Oceania, and Asia. The list of countries included in the database in January 2023 is shown in Figure 1, where the historical depth of the database is also illustrated: for each population, the period covered runs from the earliest to the most recent year, using available annual vital statistics on births by multiplicity. For example, the oldest series are those for Sweden and Denmark, with uninterrupted data since 1815 and 1850, respectively. For some countries, interruptions are observed in certain years due to limitations in data availability. An effort is made to document those periods as much as possible in the country-specific metadata files.

For any given country, the format of the original sources ranges from printed volumes of historical statistics to downloadable data files. For example, historical data for the 19th and most of the 20th century come from printed tables in the statistical yearbooks published by each country’s national statistics office. Much information from those sources was manually digitized, as the data available from the websites of the national statistics offices usually cover a recent period (e.g., from the 1970s or even the 1990s onwards). The latter type of data is usually in a format already suitable for computational analysis, such as .csv or .xlsx files. For a few countries, data were obtained upon request from national statistics offices because their websites provide only limited
information on multiple births. Complete references for each data source can be found in the country-specific metadata files.

**Figure 1:** Coverage of the HMBD according to the availability of twinning rates, by country

Source: Human Multiple Births Database - HMBD (2022).
Note: For England and Wales and Italy, the series of estimates on multiple rates (i.e., the rate of all multiple births, including twins) are more complete than those on twinning rates, covering most years from 1938 to 2019 (England and Wales) and from 1868 to 2019 (Italy).

### 2.2 Data processing and availability

For each country included, the HMBD aims to offer documented, long annual data series of the indicators shown in Table 1. To this end, the input data available for each country and period are digitized, verified, corrected, and used to calculate the missing variables if necessary. This section describes the procedure to obtain HMBD’s pooled data file (i.e., the final database that includes all countries) as well as the documentation files.

To obtain HMBD’s pooled data file, the first data processing step consists in entering the available information on births by multiplicity from the official sources (e.g., annual yearbooks or data files provided by national statistics offices) into a spreadsheet
with a column for each variable shown in Table 1. This is the input data file, which is available for each country in the HMBD website.

Table 1: Variables included in the Human Multiple Births Database (HMBD)

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country</strong></td>
<td>Country name (e.g., Australia)</td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td>Source name (abbreviation or code; e.g., “ABS” for Australian Bureau of Statistics)</td>
</tr>
<tr>
<td><strong>Year</strong></td>
<td>Reference year of the data</td>
</tr>
<tr>
<td><strong>Stillbirths</strong></td>
<td>Indicates whether stillborn children are counted:</td>
</tr>
<tr>
<td></td>
<td>0 = No (stillbirths are not included);</td>
</tr>
<tr>
<td></td>
<td>1 = Yes (stillbirths are included);</td>
</tr>
<tr>
<td></td>
<td>2 = Mixed (stillbirths are included in some cases only);</td>
</tr>
<tr>
<td></td>
<td>99 = Information not available.</td>
</tr>
<tr>
<td><strong>Singletons</strong></td>
<td>Number of single deliveries</td>
</tr>
<tr>
<td><strong>Twin_deliveries</strong></td>
<td>Number of twin deliveries</td>
</tr>
<tr>
<td><strong>Triplet_deliveries</strong></td>
<td>Number of triplet deliveries</td>
</tr>
<tr>
<td><strong>Quadruplet_plus_deliveries</strong></td>
<td>Total number of deliveries involving four or more children</td>
</tr>
<tr>
<td><strong>Multiple_deliveries</strong></td>
<td>Total number of multiple deliveries (i.e., the sum of twin, triplet, and quadruplet + deliveries)</td>
</tr>
<tr>
<td><strong>Multiple_children</strong></td>
<td>Number of children born from multiple deliveries (twins, triplets, etc.)</td>
</tr>
<tr>
<td><strong>Total_deliveries</strong></td>
<td>Total deliveries (number of single and multiple deliveries combined)</td>
</tr>
<tr>
<td><strong>Twinning_rate</strong></td>
<td>Number of twin deliveries / total deliveries, per 1,000.</td>
</tr>
<tr>
<td><strong>Multiple_rate</strong></td>
<td>Number of multiple deliveries / total deliveries, per 1,000.</td>
</tr>
<tr>
<td><strong>Footnotes</strong></td>
<td>Indicates the footnote number(s) for the year concerned in the country-metadata file.</td>
</tr>
</tbody>
</table>

Note: * For some countries during specific periods, the column `Triplet_deliveries` also includes the number of quadruplets and more when it is impossible to distinguish the deliveries by number of children (among the deliveries involving at least three children) in the original data sources. These cases are indicated in the metadata files. ** At the time of writing this article, Footnotes is a recently added column, only available in the input data file of certain countries. Once this new feature is implemented for all countries in the database, it will be included in the pooled data file of the HMBD.

Next, the consistency of the input data is verified. Basic data quality checks consist in verifying whether the sum of the number of births by multiplicity equals the total number of births reported in a given country and year. Another typical verification involves calculating the number of children born in a given year based on data on deliveries by multiplicity (for example, two children are born from one twin delivery) and comparing the resulting number with the total number of children born, as reported by the national statistics office. The treatment of stillbirths in the statistics on births by multiplicity (i.e., if and how they are counted) is taken into consideration for all verifications and calculations. Usually, the consistency of the input data is confirmed by these verifications. However, in the rare instances when substantial discrepancies are detected – i.e., inconsistencies that are large enough to have a substantial impact on other calculated variables, especially the twinning rate and the multiple rate – data for the years involved are removed from HMBD’s pooled data file, but they are still available in the country-specific input data file. In addition, years with implausible data – that is, with too high or too low twinning or multiple rates – are identified with the ‘tsouliers’ function available in the forecast package for R (Hyndman et al. 2022; Hyndman and Khandakar
2008). Explanations for those years are available in the metadata files wherever possible. When such issues are not documented in the websites of national statistics offices, we try to obtain information by contacting staff working in those offices.

Given the heterogeneity in the input data’s level of detail between countries and over time, the next step is to compute some variables to obtain as much information as possible for all features shown in Table 1. For example, if only the number of twin children born in a given year is reported in the official sources, the number of twin deliveries is calculated by dividing that number by two. The calculations are carried out with separate R-codes by country; i.e., computer codes in R programming language, which contain the steps and calculations that we performed to obtain the information for the variables provided in the database from the input data of each country.

Finally, once all relevant input data are entered and verified and all necessary calculations are performed, a single, pooled data file is produced. This final data file and all country-specific metadata, input data files, and R-codes are available for downloading at the HMBD website. Furthermore, rapid visualization of the final data file is facilitated through interactive, user-friendly web applications, built with the R Shiny package (Chang et al. 2021) and available on the HMBD website. One of the two web applications available at the time of writing this article can be used to explore all the data for a given country in the HMBD (see an example in Figure 2); the other one facilitates country comparisons of the twinning and the multiple rates.

The process described above is repeated at each update, which occurs annually for each country included in the database (see Appendix 1). At each update we extend the series of each country by adding the data for the most recent year available and verify whether official data for previous years has been corrected by the corresponding national statistics office. Such revisions are included in our files, as we change the old information with the revised data.

In addition to updated, long series of data, the HMBD aims to provide complete documentation for each country’s data. In order to do so, we search for information about the input data on the websites of the national statistics offices – including searching through statistical yearbooks and annual reports, one by one – and we contact them directly whenever we have questions about the data (for example, for years with problematic data). We synthetize that information in country-specific metadata files and whenever possible we contact experts to validate the information provided in those documents. Finally, we keep the metadata files up to date, as we add further details when new input data become available. We also occasionally introduce other modifications in order to enhance or enrich the available documentation. For example, at the time of writing this paper we are implementing a new format for the country-specific metadata files in order to provide more structured documentation, including separate sections for data coverage and completeness, definitions and the treatment of births by vitality in the
Data on multiple deliveries by certain characteristics (such as the sex composition of new-borns, mother’s age, etc.) may also be available for some countries over specific periods. As this detailed information is available only for a minority of countries, it is not included in the annual series provided by the HMBD. Instead, a list of sources where such information can be found is provided as supplementary material on the HMBD website. The website also provides a glossary of relevant terms regarding demographic data on multiple births.
3. Challenges with data on multiple births

Given possible differences in birth registration practices, data collection, and processing, comparing data on multiple births between countries and/or over time can be challenging. For example, the definitions used and the population covered may vary across sources. The HMBD provides researchers with metadata to assess when such comparisons are possible. The most critical of these challenges, as identified during the construction of the HMBD, are presented in this section.

A fundamental aspect to consider, and a key potential difficulty, relates to the measurement units provided in the sources (i.e., what is counted in the statistics on multiple births), as these may vary between countries and over time. In some sources the word ‘birth’ refers to children born, while in others it refers to deliveries, and certain statistics leave some doubt about the event recorded. This may lead to confusion. The twinning rate can vary by a factor of two according to the definition used. The definition most commonly used, and the one adopted in the HMBD, is the proportion of all deliveries that are deliveries of twins. As shown in Table 1, the HMBD focuses on the annual number of deliveries and provides their distribution by multiplicity. The total number of children born from multiple deliveries is also provided. To understand the key terms used in the database, the website provides a glossary of relevant concepts along with the data and metadata files.

Data comparability may also be affected by the inclusion (or exclusion) of the stillbirths in the statistics. For example, for a given country, stillbirths may not be included in the historical data (e.g., for the late 19th and early 20th centuries), but they may be for more recent years. In cases of twin births involving one live birth and one stillbirth the delivery may be counted as a single or twin delivery, depending on the definitions used and the reporting practices in each country. If both children are stillborn, the delivery may or may not be counted. Moreover, the definition of what counts as a stillbirth may vary between countries as well as over time within the same country. For example, in England and Wales, the Stillbirth (Definition) Act of 1 October 1992 changed the definition of a stillbirth, moving the threshold of completed weeks of gestation from 28 to 24 weeks. All these issues are documented, as much as possible, in the country-specific metadata files. These documents contain relevant notes and definitions regarding the treatment of stillbirths in the statistics – an important issue because stillbirths are more common in pregnancies involving multiple children. Unfortunately, the treatment of stillbirths in the statistics on multiple births remains unknown for certain countries and periods, as some sources do not provide sufficient information. In the database, the variable ‘Stillbirths’ indicates whether these are included in the data for a given country and year, including a specific code for cases where that treatment is unknown (see Table 1).
Other events or historical processes that may compromise data comparability, such as territorial changes and population coverage, are also explained in each country’s metadata file.

4. Database use: Example of analysis using HMBD data

Using a precursor of the HMBD, Pison, Monden, and Smits (2015) show that in developed countries for which long series of statistical information are available the twinning rate remained at roughly the same level during the late 19th century and the first half of the 20th century. From the 1950s to the mid-1970s it decreased in nearly all countries, regardless of the previous level, reaching a historical low. Then, from the mid-1970s an unprecedented and rapid increase occurred, with figures doubling over the next 40 years in many countries. Figure 3 shows this pattern, using updated data from the HMBD.

**Figure 3: Change in the twinning rates in selected countries**

Covering more world regions and using a mix of data sources including the HMBD, Monden, Pison, and Smits (2021) find that since the 1980s the global twinning rate has increased by a third, from 9.1 to 12.0 twin deliveries per 1,000 deliveries. In 2010–2015
the number of twin deliveries was higher than ever, both at the world level and for all global regions except South America, where it declined slightly. While global deliveries have increased by only 8%, the number of twin deliveries has increased by 42%.

As mentioned previously, two factors are mainly responsible for this increase: the delay in childbearing age and the rise in medically assisted reproduction (MAR). MAR includes complex techniques such as artificial reproductive technologies (ART), e.g., in vitro fertilization and intracytoplasmic sperm injection, but it also includes simpler techniques such as ovarian stimulation. These treatments increase the likelihood of twin births. MAR, which has increased substantially since the 1970s (Wyns et al. 2020), explains between 22% and 87% of the total change over the period 1970–2005, depending on the country (Pison, Monden, and Smits 2015). Delayed childbearing – which has been observed in many countries over the last decades – has also contributed to increasing the twinning rate, as the likelihood of twin pregnancy increases with mother’s age (over the age range 15–37 years), even in the absence of MAR (Bulmer 1970).

Delayed childbearing and MAR are the major factors responsible for the sharp increase in twinning rates since the 1970s. While other factors may also play a role their contribution is probably small, and for some of them the existing evidence is inconclusive, as there are contradictory results. Birth order, genetic predisposition in families, differences across regions and ethnic groups, and oral contraceptive use have been found to influence the frequency of twin pregnancies (Hoekstra et al. 2008; Harlap 1979; Bulmer 1970; Heuser 1967). However, with the exception of parity, which has changed considerably as a result of birth control, there is no reason to expect their influence to have changed much in recent decades. The parity effect (i.e., more twins at higher parity) may have played some role in the trends in twinning rates during the demographic transition, as fertility levels have decreased and hence fewer higher-parity children have been born (Pison and Couvert 2004). However, in the countries considered by the studies cited above, this decrease was almost completed at the end of the 1960s and the distribution of births by parity has not changed much since the 1970s (Frejka et al. 2008). The major remaining factors responsible for the increase in twinning rates since the 1970s are therefore the increase in mother’s age at childbearing and the expansion of MAR.

The increase in twinning rates due to MAR has raised concerns among governments and medical authorities, leading to changes in MAR regulations and practices in most developed countries. New criteria for evaluating the success of ART have been developed, focusing on successful deliveries of singleton live births. The mean number of embryos transferred has been reduced (Wyns et al. 2020). These changes are probably responsible for the finding that in about a quarter of the countries studied the twinning rates plateaued in the early 2000s and decreased afterwards (Pison, Monden, and Smits
Recently this development has occurred in more and more countries, as revealed by the updated series of the HMBD. Furthermore, as shown in Figure 3, countries differ significantly in the timing and levels of the maximum values reached. Ongoing specific data analysis projects using the HMBD include documenting recent levels and trends in twinning rates in different countries and attempting to disentangle the effects of various factors on changes in the twinning rate.

5. Future prospects

The HMBD website is hosted at the French Institute for Demographic Studies (INED) through its DataLab, an infrastructure for developing research databases. Since its creation in 2000 the development of the database has been supported by INED and the French Museum of Natural History. These two public French institutions intend to continue their support for the HMBD as part of their duty to maintain, enlarge, and update data collections useful for the research community and the public.

As the HMBD gathers annual vital statistics data on multiple births, data series are updated every year for each country when its national statistics office publishes new data for the most recent year. Keeping these series updated and readily available through the HMBD will help to track and compare changes in the rate of twin and multiple births in different countries over the years ahead. The aim is to understand the factors behind these changes – a condition for better managing this important public health issue, as explained in the previous section. In addition to data updates, we occasionally revise the metadata files for each country in order to add further information about the data or to convey that information in a more suitable way for users. For example, at the time of writing this paper we are implementing changes related to the documentation of the data, as we are updating and re-structuring the country-specific metadata files, as described at the end of section 2.2.

6. Acknowledgements

The HMBD would not be possible without the efforts of national statistics offices to provide data on multiple births. Staff members in some of these offices were particularly helpful, providing data unavailable online and clarifying relevant aspects of metadata. We are grateful for their help. We also thank researchers in various countries who helped us find historical data. The HMBD is funded by the French National Research Agency (Agence Nationale de la Recherche, ANR) and is part of JUMEAUX, a joint project of the French Museum of Natural History (Muséum National d’Histoire Naturelle) and the
French Institute for Demographic Studies (Institut National d’Etudes Démographiques, INED).
References


Pison, G., Monden, C., and Smits, J. (2015). Twinning rates in developed countries:

diabetes in women with twin pregnancies. Journal of Maternal-Fetal and

Wyns, C., Bergh, C., Calhaz-Jorge, C., De Geyter, C., Kupka, M.S., Motrenko, T.,
Rugescu, I., Smeenk, J., Tandler-Schneider, A., Vidakovic, S., and Goossens, V.
(2020). ART in Europe, 2016: Results generated from European registries by
Appendix

Figure A-1: Process for including data updates for any given country in the Human Multiple Births Database