Comparative Population Studies
D1-D9 (Date of release: 04.08.2016)

The 2011 Census Model in Germany

Sabine Bechtold

1 Overview of the German Census 2011

The concept of traditional censuses, as they were held in Germany before the recent 2011 census, is quite easy to understand. For every person living in Germany at the reference day of the census, a paper questionnaire was used to gather information. The data collection was conducted face-to-face by an enumerator, or respondents returned a filled-in questionnaire which had been delivered by an enumerator. Taking part in the census was legally mandatory. The result was a data set with one record for each person containing all census variables. As any statistical survey, these censuses were not without errors. However, the errors were restricted to systematic errors. There were no random sampling errors, which only occur among sample surveys.

In contrast to this direct approach of traditional complete enumerations, the 2011 census used different data sources and methods to collect information on persons and households as well as on buildings and dwellings. The key concept of the 2011 census was based on the idea of using the demographic basis information retrieved from the decentralised population registers and complete – and where necessary, correct – this data by merging it with information from other registers and mandatory primary surveys.[1] These surveys were designed as sample surveys or as complete enumerations and were conducted as postal or oral interviews. By merging different data sources and methods of automatic data generation, a distinct data record containing all required census information was created for each person, each household and each building with dwellings. The results for buildings and dwellings are subject to systematic errors only, which are common among all statistical surveys. They are not subject to the random sampling errors of sample surveys. In the case of data on individuals, this only applies to the demographic basis information of selected subpopulations only for which data was collected by complete enumeration.

[1] A detailed description of the concept of the 2011 census is available in German in *Statistische Ämter des Bundes und der Länder* 2015.

URL: www.comparativepopulationstudies.de
DOI: 10.12765/CPoS-2016-07en
URN: urn:nbn:de:bib-cpos-2016-07en2

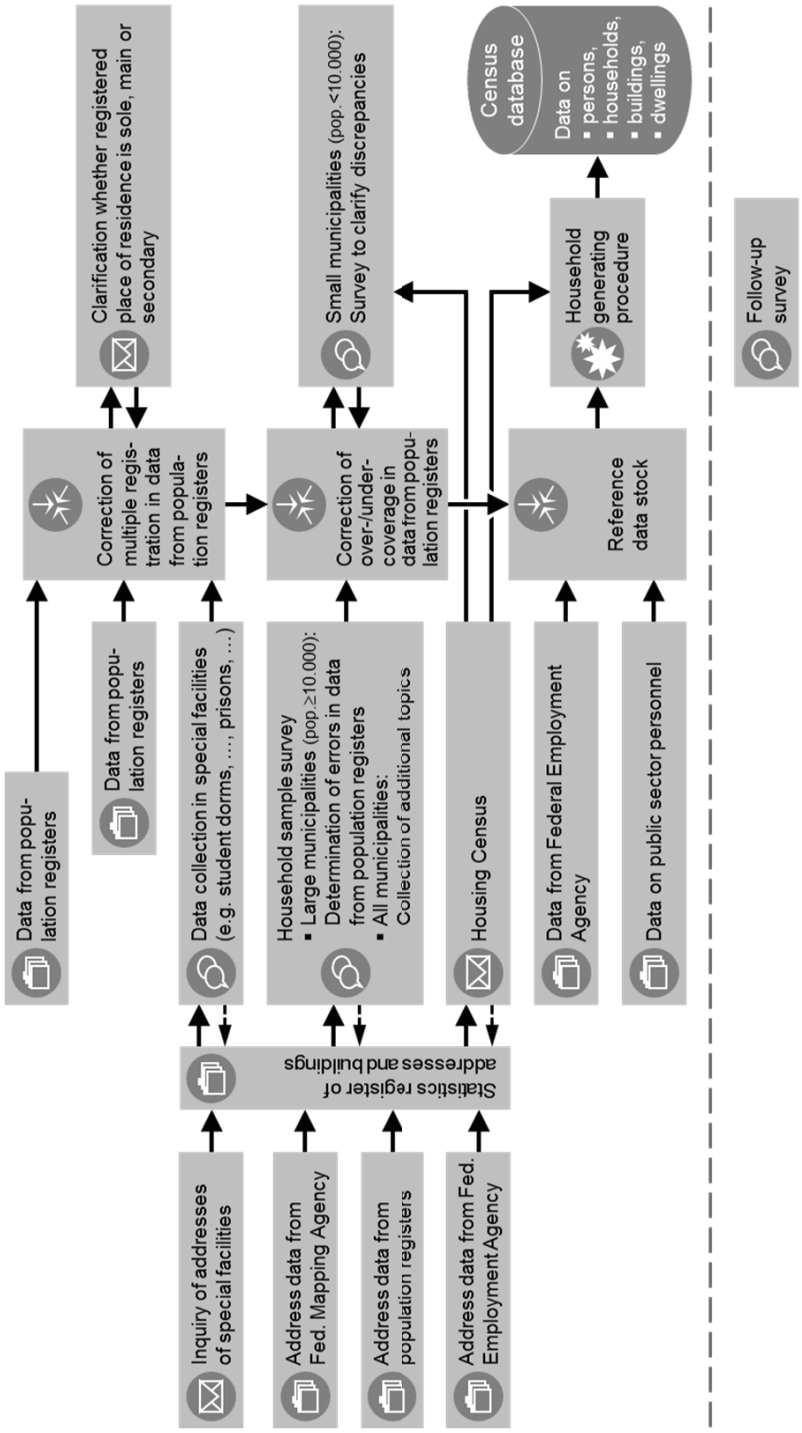

Fig. 1: 2011 Census Model

Source: own design

In order to merge the data of the different parts of the census data collection, a basic register was established, containing a list of all addresses where dwellings existed at the census reference day. This address and building register (AGR) was the key link for all data collections during the census. It was also used as the statistical population for the sampling procedure of private households and for the housing census. The most comprehensive data for the AGR was derived from the Federal Mapping Agency and included residential as well as non-residential buildings. Address data from the decentralised population registers and the register of the Federal Employment Agency were also used. To prevent a systematic under-coverage of the population all buildings relevant for the census had to be included in the AGR. Therefore all addresses only existing in one data source were cross-checked to confirm whether they contained dwellings or not. The AGR was furthermore updated with additional data and information gathered during the mandatory primary surveys (*Kleber et al.* 2009).

Data from the decentralised population registers maintained by the municipalities at the date of the census reference day (9. May 2011) were the basis for the calculation of the census population figure for each municipality and also the demographic structure of the population in private households by sex, age, legal marital status and citizenship. The data from the population registers were collected at the census reference day and were updated three months after the census reference day in order to cover delayed register entries. These register data were merged to a nationwide data set and it was subsequently tested whether people were registered at more than one sole or main place of residence on the census reference day. If such cases were identified in large municipalities (with at least 10.000 inhabitants), they were automatically corrected by using the most current information. Multiple residences in small municipalities (with less than 10.000 inhabitants) were investigated using a postal inquiry. The same applies to cases where a person was registered at a secondary place of residence only (*Diehl* 2012).

For persons living in special facilities, e.g. a communal accommodation, care institution, dormitory or similar types of housing, census information was collected using a complete enumeration because fluctuation and missing registrations for this sub-population lead to high rates of error in the population registers. Addresses carrying stigmatizing information, e.g. in the case of psychiatric hospitals or prisons ("confidential special facilities"), were distinguished from non-confidential special facilities, e.g. student dormitories. In confidential special facilities, the privacy of data collection was secured by a special procedure. Persons living at special facilites were included in the test on multiple residences described above.

The household sample survey of almost 10 percent of the population was used to ensure the quality of the register data in large municipalities which was already corrected for multiple residences. For the calculation of the population of large municipalities, the level of error of the population registers (over- and under-coverage) was taken into account. Further, the sample survey was designed to collect additional information on the population in large and small municipalities which could not be gained from registers. This applies to the complete enumeration as well,

which was held in non-confidential special facilities.[2] The additional information collected included the educational attainment, migration background, religion respectively religious orientation and employment. The regional differentiation of this information is restricted to large municipalities and NUTS-3-regions only.

In order to use the household sample survey to identify over-coverage as well as under-coverage of the population, a random sample of addresses was used. For all people actually living at these addresses the residential status was compared with the respective information contained in the population registers. Sampling frame were all addresses in the AGR except those with confidential special facilities.

The sample design was developed to ensure that the population figures of large municipalities meet a 1 percent error margin target at a 95 percent confidence level. In order to optimise the sample design, a stratified sample was used and the sample size dedicated by the census law was distributed among the stratified subsamples to attain the highest possible precision for the calculation of the population size for each municipality. The information of the assumed amount of over- and under-coverage in the population registers by municipality, which is of great importance for the precision achieved by the 2011 census and therefore used for the optimisation of the sampling process, was based on a census test held in 2001 (*Statistische Ämter des Bundes und der Länder* 2004). However the actual amount of over- and under-reporting by municipality could only be obtained in the census itself. The method applied to optimise the sampling process was dedicated individually to each municipality and the sample size ranged between 2.1 percent and 45.6 percent and differed significantly even for municipalities of a similar size.

Register data of the Federal Employment Agency on employees subject to social insurance contributions and register data on the personnel of public employers were used to supplement the individual demographic information of the population registers, the survey of private households and the survey of addresses with non-confidential special facilities. Together with the AGR, this information constituted the reference data set (RDB) (*Hirner/Stiglmayr* 2013). The RDB allowed the validation of the information contained in all data sources of the census and was used to assign individual records to a geographical reference, such as administrative units or a countrywide grid of 100 by 100 meters.

Merging data sets from different sources for individual persons was one of the great challenges of the 2011 census, because it had to be accomplished without an existing personal identification number. Individual and address-based information such as name, sex, date of birth, municipal code, post code, street name, and house number were used to link respective records of different data sets. Only 1 percent of all records couldn't be linked automatically and had to be matched manually.

---

2 The sample design and extrapolation model for the household sample survey was developed by an external research project at the University of Trier and GESIS – Leibniz-Institute for Social Science in Mannheim (see *Münnich et al*. 2012). The methods developed for the 2011 census are described in *Berg/Bihler* 2011 and *Berg/Bihler* 2014a.

Information on the number and structure of buildings and dwellings were collected by a postal census held among all property owners. Information on the type of building, year of construction, number of residential dwellings and heating system was collected. For dwellings, additional information on the size and number of rooms and the use of the dwellings was collected, among others (*Statistische Ämter des Bundes und der Länder* 2014). In small municipalities, the housing census was also used to test and – where necessary – to make additional inquiries to correct the data obtained from population registers. For single-family homes and other addresses with one occupied dwelling, a positive correlation was predicted between the additional effort of the inquiry and the effect of the potential correction on the precision of the results. The impact of these corrections on the quality of the information obtained from the population registers was expected to add up to the level of precision gained from the random sample in large municipalities.

To obtain information on how persons live in households and on their housing conditions, an automatic procedure was used to combine the information of persons with statistically generated household units. Therefore for each address, population register data, which included information on the relation of family members, were merged with information from the housing census.

In order to evaluate the population of large municipalities, a follow-up survey of about 5 percent of the addresses of the household sample survey was held by enumerators (*Klink*/*Bihler* 2015).

2 Distinct features of the census data

All 2011 census data were collected by mandatory surveys. Only one question on the religious orientation was optional. Nevertheless, data had to be validated and corrected where necessary, e.g. if they were implausible or missing completely. To this end, cold deck and hot deck as well as deterministic imputation methods were used. For the housing census, 31 percent of the questionnaires were plausible and complete. The necessary imputation was especially high for the indicator of ownership, which had to be imputed for 18.5 percent of all buildings and 45.3 percent of all dwellings.[3] All other questions on housing units had rates of imputation of less than 16 percent (*Grundwald*/*Krause* 2014). The questionnaires of the household sample survey were plausible at a level of approximately 50 percent. Among the different thematic sections, questions on the activity status had the highest rate of implausible records, reaching 25 percent (*Statistisches Bundesamt* 2016: 31). The optional question on the religious orientation was answered by only 48 percent of those respondents not belonging to an officially recognized religious community. Because of the high item-nonresponse rate and the estimated large error margin, the German Federal Statistical Office decided not to publish results on this variable.

[3] The high value is due to an error of the questionnaire design. The question on ownership was misleading, but the response to this question does not affect the quality of the survey as a whole.

The census population figure for large municipalities originates from two different parts. The population at addresses with special facilities was calculated based on a complete enumeration (enumeration section). For all other addresses, the population of the population registers, corrected for multiple residences, was adjusted by the rate of over- and under-coverage extrapolated from the random household sample survey (adjustment section) (*Berg/Bihler* 2014b). Both sections add up to the population figure of a municipality. The rate of over- and under-coverage for each municipality was derived from the estimation of the population contained in the population registers at the census reference day and the number of persons who were correctly registered. This method was deemed preferable to a direct estimation of the rate of over- and under-coverage because the estimation error was expected to be lower. Part of the extrapolation procedure was a test if the precision of the calculated population size meets the intended target. In fact, the target could only be met in 37 percent of the municipalities because the assumptions underlying the sample design were too optimistic (*Statistisches Bundesamt* 2016: 7).

For small municipalities, the population figure of the 2011 census was calculated by counting the persons living at addresses with special facilities, which were enumerated completely, and the population register data which was corrected for cases of over- and under-coverage. These corrections were based on the above mentioned inquiry investigating implausible cases in small municipalities.

The correct interpretation of the 2011 census data on individuals requires considering the data sources from which the information was retrieved. Those results stemming from the household sample survey, and not from register data, are not available for persons living in confidential special facilities. In those cases where data based on the random household sample survey were compiled together with demographic basic variables, the latter were also derived from the random household sample survey and the totals can therefore differ from the demographic basic results based on register data. The data on employment were partly collected from registers and partly from the household sample survey and from the data collection at addresses with non-confidential special facilities. The data on activity and occupational status and the classification of economic activities can be counted for employees subject to social insurance contributions and for the personnel of public employers, and have to be extrapolated for all others. In cases where these data were compiled together with data from the household sample survey, they were also estimated from the household sample survey. This might result in differences compared to the register-based results (*Sedmihradsky et al.* 2012).

Data based on the complete enumeration or registers can be counted directly, data from the household sample survey have to be extrapolated. Therefore, different procedures to safeguard the confidentiality of the data are required. The data protection of the results of the housing census, of the automatic household generation procedure and data from the population registers as well as data collected at the addresses with non-confidential special facilities was guaranteed using the SAFE procedure, which modifies the reported data (*Gießing et al.* 2014). For all municipalities and for all districts of Berlin and Hamburg it was assured that the cell frequency of each combination of individual variable values and each combination of

variable values from the housing census as well as the automatic household generation was three or more. In contrast to this procedure, the total population figure of the municipalities was calculated without modifying the data. The ratios published by the statistical offices were also calculated based on unmodified population data. In these cases, the privacy of the data is protected by rounding the division results.

For the extrapolated results, no separate procedure of disclosure control was used. If results were subject to a relative standard error of more than 15 percent, they were regarded as unreliable and were not published. In cases where results were compiled from both counted as well as extrapolated data, the respective contribution of both parts to the final result was considered for the decision on whether to publish the data or omit it. All results derived, in part or fully, from extrapolated data were rounded to a multiple of 10 to indicate that the figure is not a precise value. The combination of sampling errors, omission of data cells and rounding guarantees statistical disclosure.

The demographic results of the 2011 census for each municipality, differentiated by sex, age, marital status and citizenship, are the basis for population estimates used until the next census results are available. They are prepared by updating the census population data with information from vital and migration statistics. The census population figures and their differentiation by demographic variables that feature as a basis for the population estimates were calculated by a specific procedure which might, for particular municipalities, result in demographic structures differing from the general census results. However, the total population size is not affected by this potential deviation. Using a different procedure was necessary because the census extrapolation method was optimised to attain the highest possible precision for the population size. In some municipalities the procedure resulted in suspicious figures for certain combinations of age and sex. Without an adjustment this would not only negatively affect the future population estimates until the next census but would also have a negative impact on the calculation of demographic indicators such as birth and death rates (*Statistisches Bundesamt* 2015).

References

Berg, Andreas; Bihler, Wolf 2011: Das Stichprobendesign der Haushaltsstichprobe des Zensus 2011. In: Wirtschaft und Statistik 04/2011: 317-328 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaApril11.pdf?__blob=publicationFile, 25.07.2016].

Berg, Andreas; Bihler, Wolf 2014a: Der Auswahlplan für die Ziehung der Neuzugänge der Haushaltsstichprobe des Zensus 2011. In: Wirtschaft und Statistik 03/2014: 151-154 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaMaerz2014.pdf?__blob=publicationFile, 25.07.2016].

Berg, Andreas; Bihler, Wolf 2014b: Das Hochrechnungsverfahren zur Ermittlung der Einwohnerzahl im Zensus 2011. In: Wirtschaft und Statistik 04/2014: 229-235 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaApril2014.pdf?__blob=publicationFile, 25.07.2016].

Diehl, Eva-Maria 2012: Methoden der Mehrfachfallprüfung im Zensus 2011. In: Wirtschaft und Statistik 06/2012: 473-484 [https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaetze_Archiv/2012_06_Destatis_Methoden_der_Mehrfachfallpruefung_im_Zensus_2011.pdf?__blob=publicationFile&v=14, 25.07.2016].

Gießing, Sarah et al. 2014: Geheimhaltung beim Zensus 2011. In: Wirtschaft und Statistik 11/2014: 641-647 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaNovember2014.pdf?__blob=publicationFile, 25.07.2016].

Grundwald, Sven; Krause, Anja 2014: Umgang mit fehlenden Angaben in der Gebäude- und Wohnungszählung 2011. In: Wirtschaft und Statistik 08/2014: 437-449 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaAugust2014.pdf?__blob=publicationFile, 25.07.2016].

Hirner, Stephanie; Stiglmayr, Susanne 2013: Der Referenzdatenbestand im Zensus 2011. In: Wirtschaft und Statistik 01/2013: 30-41 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaJanuar2013.pdf?__blob=publicationFile, 25.07.2016].

Kleber, Birgit et al. 2009: Aufbau des Anschriften- und Gebäuderegisters für den Zensus 2011. In: Wirtschaft und Statistik 07/2009: 629-640 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaJuli09.pdf?__blob=publicationFile, 25.07.2016].

Klink, Steffen, Bihler, Wolf 2015: Die Wiederholungsbefragung beim Zensus 2011. In: Wirtschaft und Statistik 02/2015: 42-54 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/2015/02/Wista_2_2015.pdf?__blob=publicationFile, 25.07.2016].

Münnich, Ralf; Gabler, Siegfried u.a. 2012: Stichprobenoptimierung und Schätzung in Zensus 2011. Statistik und Wissenschaft Band 21. Wiesbaden: Statistischen Bundesamt [https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band21_StichprobenoptimierungZensus1030821129004.pdf?__blob=publicationFile, 25.07.2016].

Sedmihradsky, Dirk; Eisenmenger, Matthias; Burmeister, Friederike 2012: Erhebung, Aufbereitung und Zusammenführung der Erwerbsregisterdaten im Zensus 2011. In: Wirtschaft und Statistik 12/2012: 1052-1064 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaDezember12.pdf?__blob=publicationFile, 25.07.2016].

Statistische Ämter des Bundes und der Länder 2004: Ergebnisse des Zensustests. In: Wirtschaft und Statistik 08/2004: 813-833 [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaAugust04.pdf?__blob=publicationFile, 25.07.2016].

Statistische Ämter des Bundes und der Länder 2014: Zensus 2011 – Gebäude und Wohnungen. Übersicht über Merkmale und Merkmalsausprägungen, Definitionen [https://www.zensus2011.de/SharedDocs/Downloads/DE/Merkmale/Merkmale_GWZ.pdf?__blob=publicationFile&v=13, 25.07.2016].

Statistische Ämter des Bundes und der Länder 2015: Zensus 2011 – Methoden und Verfahren. Wiesbaden 2015 [https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaetze_Archiv/2015_06_MethodenUndVerfahren.pdf?__blob=publicationFile&v=6, 25.07.2016].

Statistisches Bundesamt 2015: Ausgangsdaten der Bevölkerungsfortschreibung aus dem Zensus 2011 [https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsstand/DatenBevoelkerungsfortschreibungZensus.html, 25.07.2016].

S*tatistisches Bundesamt* 2016: Qualitätsbericht nach § 17 ZensG2011. Wiesbaden: Statistisches Bundesamt [https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaetze_Archiv/2016_05_Zensus2011_Qualitaetsbericht.pdf?__blob=publicationFile&v=8, 25.07.2016].

Dr. Sabine Bechtold (✉). Federal Statistical Office. Wiesbaden, Germany.
E-mail: sabine.bechtold@destatis.de
URL: https://www.destatis.de/EN/Homepage.html

Comparative Population Studies

www.comparativepopulationstudies.de

ISSN: 1869-8980 (Print) – 1869-8999 (Internet)

Published by
Prof. Dr. Norbert F. Schneider

Federal Institute for Population Research
D-65180 Wiesbaden / Germany