

**Quality Assessment of Administrative Data for Indonesian Population Census  
based on Census Test Results**

Winida Albertha<sup>\*1</sup>; Nurma Midayanti<sup>2</sup>; Putri I. Firdaus<sup>3</sup>; Ahmad Kosasih<sup>4</sup>

<sup>1</sup> BPS-Statistics Indonesia, Jakarta, Indonesia, winid@bps.go.id

<sup>2</sup> BPS-Statistics Indonesia, Jakarta, Indonesia, nurma@bps.go.id

<sup>3</sup> BPS-Statistics Indonesia, Jakarta, Indonesia, putri.firdaus@bps.go.id

<sup>4</sup> BPS-Statistics Indonesia, Jakarta, Indonesia, ahmad.kosasih@bps.go.id

**Abstract**

BPS-Statistics Indonesia will hold its seventh population census in 2020. In contrast to the traditional methods in previous censuses, it will use a combined method by making use of the administrative data obtained from the Ministry of Home Affairs combined with the field data enumeration. Individual information recorded in the administrative data will be used as the initial data source for the field enumeration as well as to ensure the coverage of census. The changes in the census method could be a step towards the register-based census in the future. Given the important role of register data in the upcoming Indonesia census, it is necessary to know the quality of the administrative data which have been recorded by the Ministry of Home Affairs. Therefore, this study aims to analyze whether there are significant differences in certain individual variables between the data from the administrative source and the census test results. Data of census test were obtained through the census test using multi-mode data collection in three selected villages in Indonesia. Statistical tests between the two sources showed that there are no significant differences for most individual variables tested. These results indicate that administrative data has a sufficient quality to be used as the initial data source for the combined method for the census in Indonesia. However, to be able to carry out the register-based census in the future, the quality of administrative data must be further improved, especially in the up-to-date aspect of the information that may change over time.

**Keywords:** combined method; administrative data; census test.

**1. Introduction**

As a part of an integrated national statistical system, the United Nations (2017) recommends all countries around the world to produce detailed population and housing statistics at least once in a 10-year period. Furthermore, population and housing census is also required by Indonesian constitution. Therefore, with BPS-Statistics as the agency responsible for official statistics, the country has conducted six population and housing censuses since 1961. Until its last census in 2010, data collection has been carried out using traditional methods. Through the full field enumeration, information on individuals and households were collected using questionnaires with door-to-door interviews by enumerators. The method is considered very complex and requires large resources to undertake the operation (United Nations, 2017).

Considering the shortcomings of traditional census, BPS-Statistics Indonesia began to look for alternative approaches in conducting future censuses. It is hoped that in the future, Indonesia can perform register-based census. Register-based census is considered to be cheaper, faster, and could reduce the respondent burden (UNECE, 2018). The register-based census method has been carried out by many countries in the world including the Netherlands, Denmark, Finland, Norway, Sweden, Austria, Slovenia, South Korea, Singapore (UNECE, 2018; Statistics Korea, 2018; Jialin & Lip; 2017). Before being able to conduct a full register-based census, BPS-Statistics Indonesia will perform the combined method in the upcoming census and make some changes in the business process. BPS-Statistics will utilize administrative data combined with field enumeration. The enumeration will be carried out

## 2020 Asia–Pacific Statistics Week

### A decade of action for the 2030 Agenda: Statistics that leaves no one and nowhere behind

15-19 JUNE 2020 | Bangkok, Thailand

using multi-mode data collection, which uses computer-assisted personal interviewing (CAPI) and computer-assisted web interviewing (CAWI) besides paper and pencil interviewing (PAPI).

Currently, Indonesia already has a population administrative system managed by the Ministry of Home Affairs. Every Indonesian resident registered in the system has a unique identity number. Although not all Indonesian residents are registered in the system, it is estimated that about 97% of the total population has been electronically recorded in the system (Kemendagri, 2019). Indonesian administrative system has implemented biometric verification technology. The digital identification system using biometric verification ensures that each person can only have one valid population identification number (World Bank, 2018). As using unique identity numbers is becoming the mandatory requirement to access almost all services and administration documents, there will be more and more residents who proactively request to be registered in the system. However, even though administrative data are available and have covered almost the entire population, the changes in individual data are not always updated in the system. Therefore, there may be some individual data in administrative records that is no longer the same as the current situation.

Due to the limitations in administrative data, the upcoming population census in Indonesia cannot be fully carried out on a registration basis. However, despite its limitation, BPS-Statistics Indonesia can still use the administrative source for the 2020 census. The administrative data will be used as a basic frame to ensure the census coverage. There is the possibility to borrow some individual information from administrative data to obtain certain population variables. However, to complete the housing and individual information that are not relevant or not available in the administrative data, field enumeration is still needed. Therefore, this combined census will be the starting point for Indonesian register-based census in the future.

Since the administrative data will be used in the upcoming census in Indonesia, it is necessary to do the census test. BPS-Statistics Indonesia has carried out the first test of combined method census in February 2019. The census test was conducted in three selected villages; one village in Jawa Barat province (Sukamiskin), and two villages in Kalimantan Selatan (Mekarsari and Indah Sari). The census test intended to ensure the coverage and to test the field operation. Furthermore, the census test was also carried out to assess the quality sufficiency of administrative data.

The assessment of administrative data is needed to test the variables that exist both in the administrative source and the data obtained from the census test. The results can be used to indicate which variables that can be used from administrative data and which variables are still needed to be collected from respondents. Therefore, this study aims to analyze whether there are significant differences for certain individual variables between the data from administrative source and the census test. The results may also suggest some aspects to be considered for the improvement, either in the enumeration mechanism or in the administrative data for the future censuses.

## 2. Methodology

This research will compare the variables from two data sources i.e. administrative data and data from census test. The administrative data was obtained from the Ministry of Home Affairs, while data from census test was obtained from the first combined census test carried out in one village in Jawa Barat province (Sukamiskin), and two villages in Kalimantan Selatan (Mekarsari and Indah Sari) in February 2019. From the census test, there were 1668 respondents whose data can be tested. 671 respondents were obtained from those who independently completed the form through CAWI, while 997 respondents were obtained through interviews by the enumerators using the CAPI mode. From the data sources, there are eight variables that can be compared between the census test results and administrative data: sex, religion, educational attainment, marital status, age, date of birth, province of residence, and province of birth.

This study used non parametric statistical analysis methods to determine whether there are differences between the eight variables obtained from administrative data and the census test. Statistical analysis of paired-sample t-tests were employed in this study. The methods are used to understand whether there were differences between the two dependent groups that are given different treatment (Siegel, 1956). The differences in treatment can be defined as the different data sources for the same individual. The data sources for the same individual were taken from administrative records and census test records. It is assumed that the data from the census test were more updated compared to the data from administrative source.

Null hypothesis that will be tested for the purposes of this study assumes that the true mean difference for the paired samples from administrative source and census test is zero. Conversely, the alternative hypothesis assumes that the true mean difference between those paired samples is not equal to zero. Because there are different data types of the eight variables, the type of paired-sample statistical tests were adjusted according to the data type of each variable. Tabel 1 shows the list of statistical tests for each variable.

**Tabel 1. Variable, Type of Data and Statistical Test**

No.	Variable	Type of Data	Statistical Test
(1)	(2)	(3)	(4)
1	Sex	Nominal with 2 categories	McNemar Test
2	Religion	Nominal with more than 2 categories	Marginal Homogeneity Test
3	Marital Status		
4	Date of Birth		
5	Province of Residence		
6	Province of Birth		
7	Educational Attainment	Ordinal, ratio, interval	Wilcoxon Test
8	Age		

**3. Result**

Based on census test, majority of respondents were female (52.1%) and Muslim (96.28%). Respondents were distributed in various age groups with most respondents being part of the group of 15-29 years old (27.2%). Meanwhile, the smallest number of respondents was in the group of 60 years old and over (6.1%). In terms of marital status, about 50.2% of respondents were married, 44.8% were never married, and 5% were divorced or widowed. Almost half of total respondents have finished senior high school or higher, while about 21.4% of respondents have not finished elementary school.

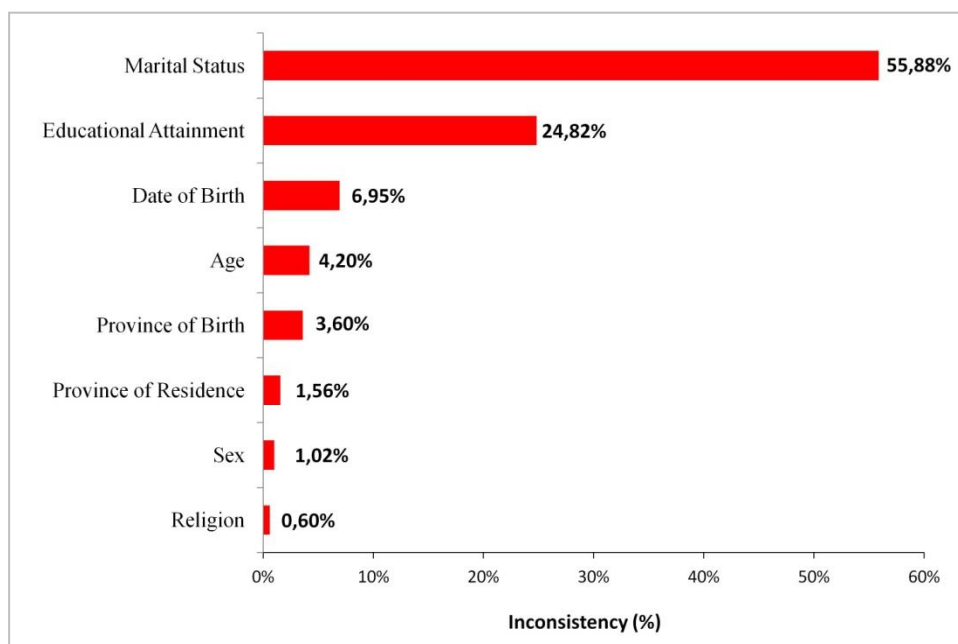
Thereafter, eight selected variables from the census test and administrative records were compared: sex, religion, marital status, date of birth, province of residence, province of birth, educational attainment, and age. The descriptive results of inconsistency for each variable are shown in Figure 1. From the results, it can be seen that the lowest inconsistency was found in variable of religion. There are only 0.6% of the total respondents that have different data on religion in census test compared to the data in administrative source. This finding suggests that religion changes in Indonesia are not very common.

Subsequently, sex variable has the second lowest inconsistency with a percentage of only 1.02%. This result indicates that administrative records for this variable were praised as being accurate. However, even the inconsistency for sex variable is considered very low, further examination is needed to ensure the correctness of the data. Although gender changes are possible, the case is considered very rare in Indonesia. Furthermore, there are also inconsistencies in the variables of the province of residence, province of birth, age, and date of

birth. However, the inconsistencies did not reach 7% of total respondents and are considered low.

In contrast with the previous variables, marital status and educational attainment have fairly large percentages of inconsistencies. There were 55.8% of respondents that have different marital status at the time of census test compared to the existing data from administrative records. Subsequently, nearly 25% of respondents have different responses on educational attainment in both sources. These results indicate that individual information on marital status and educational attainment were not updated in the administrative system.

**Figure 1. Inconsistency between the data in Census Test and Register Data**



In order to identify the statistical differences between the data from both sources, the hypothesis tests were performed on the selected variables. Table 2 shows the p-values of the test on each variable. Statistical test results for educational attainment and marital status reveal that there were statistical differences between the data from the census test and administrative records on those two variables. On the other hand, statistical tests on sex, religion, age, date of birth, province of residence, and province of birth did not show that there were any differences in the data from the two sources.

**Table 2. P-value of Each Variable**

No.	Variable	P-value
(1)	(2)	(3)
1	Sex	0.332
2	Religion	0.072
3	Educational Attainment	0.000*
4	Marital Status	0.005*
5	Age	0.357
6	Date of Birth	0.377
7	Province of Residence	0.155
8	Province of Birth	0.952

\*p< 0.05

Given that one of the data sources came from the results of census test, therefore, it was considered that errors could occur during the data collection process. There are several indications found related to this matter. Inconsistency on sex variable was explored further through identification of the respondent's name. Based on common sense and Indonesian culture, there are names that are specific to a particular gender. The results of examinations on names and sex are shown in Table 3. From the table, it can be concluded that most of inconsistencies in the sex variable were resulted from census test rather than from administrative records.

**Table 3. Sex Identification based on Individual’s Name**

No	Name	Administrative Record	Census Test Result		Identification Result
			CAPI	CAWI	
(1)	(2)	(3)	(4)	(5)	(6)
1	Meita xxxxxxxx	Female		Male	Female
2	Ahmadi xxxx xxxxxxxx	Female		Male	Male
3	Nawafiansyah xxxxx xxxxx	Male	Female		Male
4	Kevin xxxxx xxxxxxxxxxxx	Male	Female		Male
5	Syaiful xxx	Male	Female		Male
6	Winda xxxxxx xxxxxxxx	Male	Female		Female
7	M. Nashrullah	Male	Female		Male
8	Muhammad xxxxx	Male	Female		Male
9	Chaterine xxxxxxxx xxxxx	Female	Male		Female
10	Vitra xxxxxxxxxxxx	Female	Male		Female
11	Dinara xxxxxx	Female	Male		Female
12	Aqilah xxxxxx	Female	Male		Female
13	Della xxxxxxxx	Female	Male		Female
14	Migawati	Female	Male		Female
15	Nasywa xxxxxxxx xxxxx	Female	Male		Female
16	Mira xxxxxx	Female	Male		Female
17	Sariba	Female	Male		Female

Note: some names are not displayed for confidentiality reason

Examination was also conducted on the inconsistencies of educational attainment variable in both sources. There were 5% of total respondents with higher educational attainment in administrative data than in census test. Given the initial assumption in this study was that the data from census test were more updated compared to administrative data, therefore that inconsistency is an anomaly. The results of both inspections indicate that the inconsistencies in data could exist due to errors in the data collection process. The error can be caused due to misunderstandings of the concepts and definitions or due to unintentionally fill in the wrong answer or even intentionally give the wrong answers. In addition, unsuitable application interface design, both on the CAPI and CAWI could also lead to error answers.

**4. Discussion, Conclusion and Recommendations**

Register-based census using administrative source is considered cheaper, faster, and could reduce the respondent burden. Combined method census or the census that combines the use of administrative data and field enumeration is the starting point to step towards Indonesian register-based census in the future. In 2020, Indonesia will conduct its census using this

## 2020 Asia–Pacific Statistics Week

### A decade of action for the 2030 Agenda: Statistics that leaves no one and nowhere behind

15-19 JUNE 2020 | Bangkok, Thailand

method. The combined method application in the 2020 Indonesian Population Census changes the census business process compared to the traditional methods. To ensure the quality of administrative data that will be used for the census, it is necessary to assess the quality of the registration data. Therefore, BPS-Statistics Indonesia conducted the combined method census test in three selected villages in Jawa Barat (Sukamiskin), and Kalimantan Selatan (Mekarsari and Indah Sari). The census test was carried out using multi-mode data collection with gadget (CAPI) and online form (CAWI) that had never been used in previous censuses.

Evaluation of the quality of registration data was done by comparing data from administrative records with the data from census test. Descriptive and statistics test results consistently show that registration data have already a good level of accuracy on several variables such as sex, religion, age, date of birth, province of residence, and province of birth. However, there are significant differences in the variables of educational attainment and marital status.

Based on test results, it can be concluded that Indonesian administrative records can be used as the initial data source for combined method census in Indonesia. However, in order to be able to conduct fully register-based censuses in the future, the quality of administrative data must be further improved, especially on variables with a low level of accuracy, i.e. variables that are vulnerable to change over time.

Furthermore, it is important to pay attention to the quality of data filled by enumerators using CAPI and by respondents through the website. Based on the census test, it was found that there were still data errors obtained from the two modes. The errors can be caused by the lack of understanding of enumerators or respondents in certain concepts, the accidental or intentional mistakes by respondents or the insufficient user interface of the application. Therefore training of field enumerators is one of the important keys to ensure data quality. For respondents who fill out the questionnaire through the website it is necessary to provide an adequate user interface. In addition, concepts and definitions or guidelines that are easy to be understood by the public also needed to be provided.

## References

1. Jialin, C., & Lip, T. (2017). Challenge in the Development of Register-Based Population Statistics. Singapore: Statistics Singapore Newsletter.
2. Kemendagri. (9 February, 2019). Kementerian Dalam Negeri Republik Indonesia. Retrieved 1 May, 2020, from <https://www.kemendagri.go.id/berita/baca/19131/Go-Digital-Dukcapil-Siap-Wujudkan-Single-Identity-Number>
3. Siegel, S. (1956). Nonparametric Statistics for the Behavioral Science. New York: McGraw-Hill.
4. Statistics Korea. (2018). KOSTAT. Retrieved 1 May, 2020, from <http://kostat.go.kr/iwsm/download/2018/S3-4.pdf>
5. UNECE. (2018). Guidelines on the use of registers and administrative data for population and housing censuses. Geneva: UNECE.
6. United Nations. (2017). Principles and Recommendations for Population and Housing Censuses, Revision 3. New York: United Nations.
7. World Bank. (2018). Technology Landscape for Digital Identification. Washington, DC: World Bank License Creative Commons Attribution 3.0 IGO (CC BY 3.0 IGO).