総務省
MIC Ministry of Internal Affairs and Communications

総務省統計局

The 2020 Population Census
Efforts in Providing the Results Resolving the Problem of Unknown Values

SEKINO Hidemine

Deputy Director of the Population Census Division

Statistics Bureau of Japan

総務省
MIC Ministry of Internal Affairs and Communications

総務省統計局

The 2020 Population Census
Efforts in Providing the Results Resolving the Problem of Unknown Values

1. Background
2. Method of the imputation
3. Results of the imputation
4. Future development

総務省
MIC Ministry of Internal Affairs and Communications

総務省統計局

The 2020 Population Census
Efforts in Providing the Results Resolving the Problem of Unknown Values

1. Background

2. Method of the imputation

3. Results of the imputation

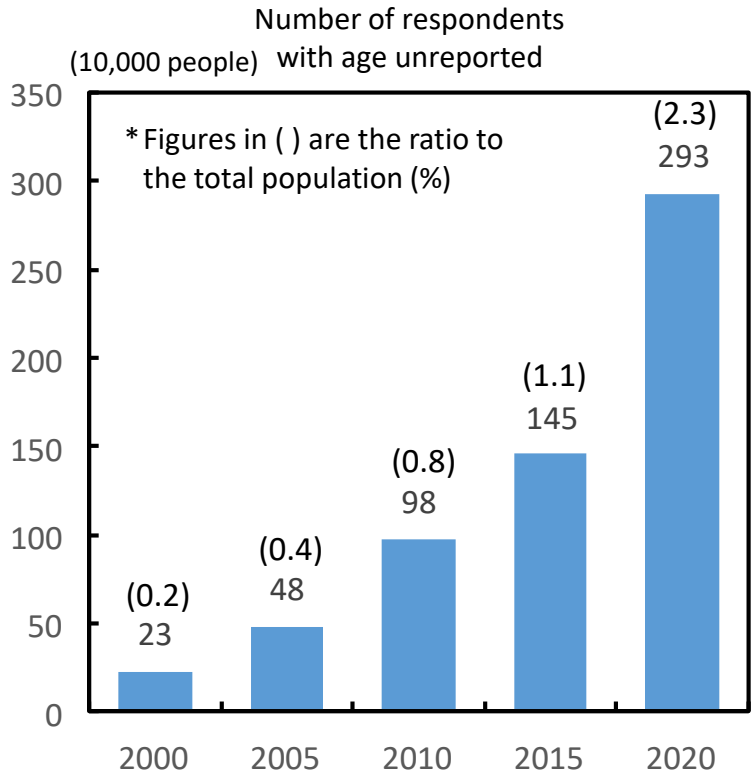
4. Future development

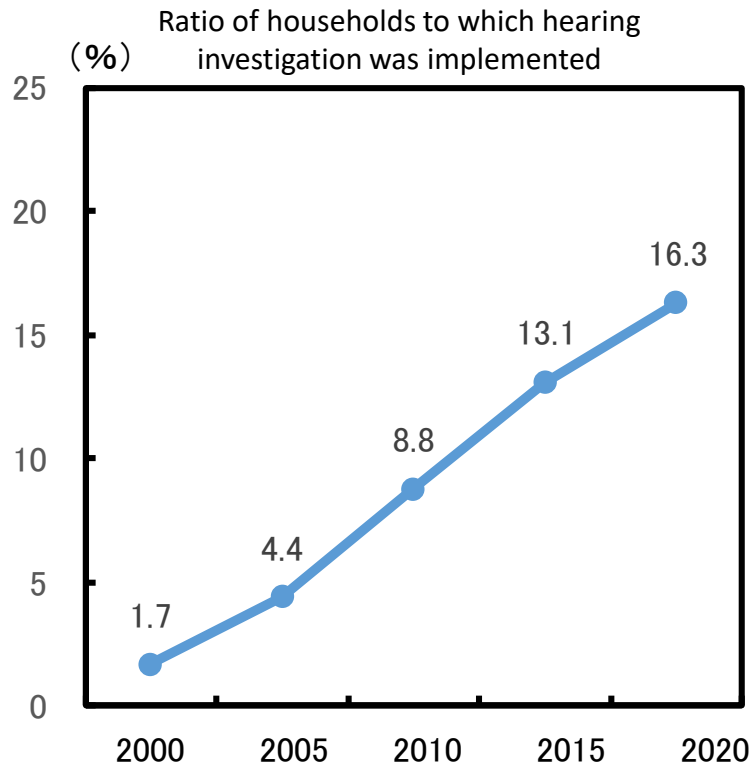
1. Background (1)

Along with the recent deterioration of the survey environment, the number of answers obtained through hearing investigation* because of inability to obtain responses directly from the target is increasing.

→ Amount of data with the detailed information on attribution "unknown" is increasing.

***As for households from which questionnaire sheets could not be collected due to absence, etc., survey is implemented by enumerators asking the "name," "sex," and the "number of household members" from neighbors, etc.**



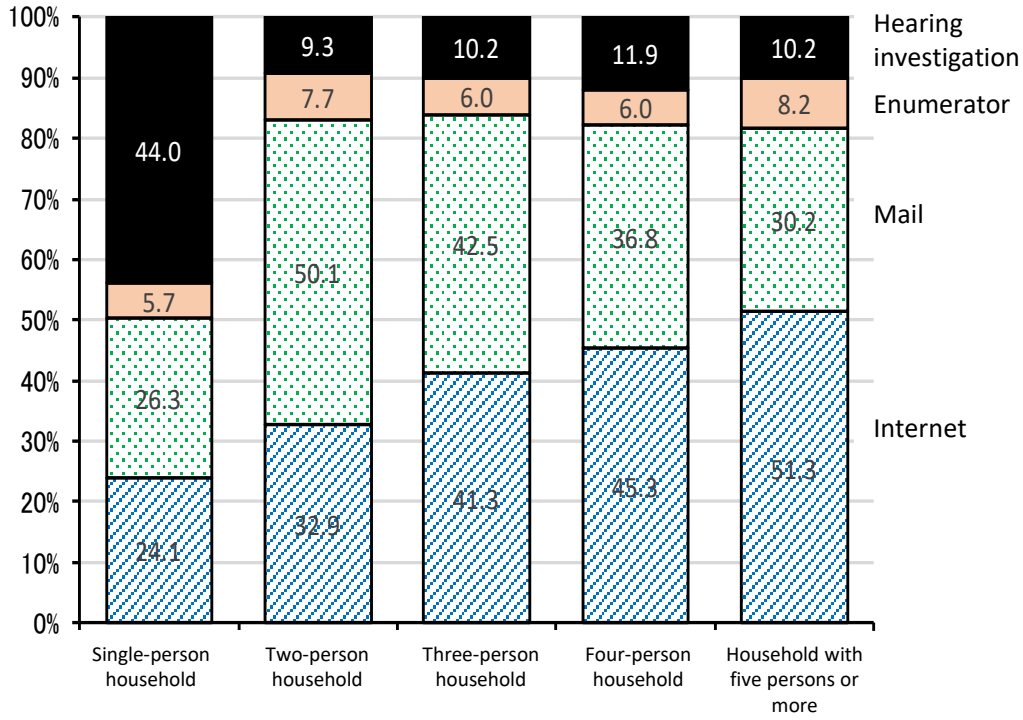


1. Background (2)

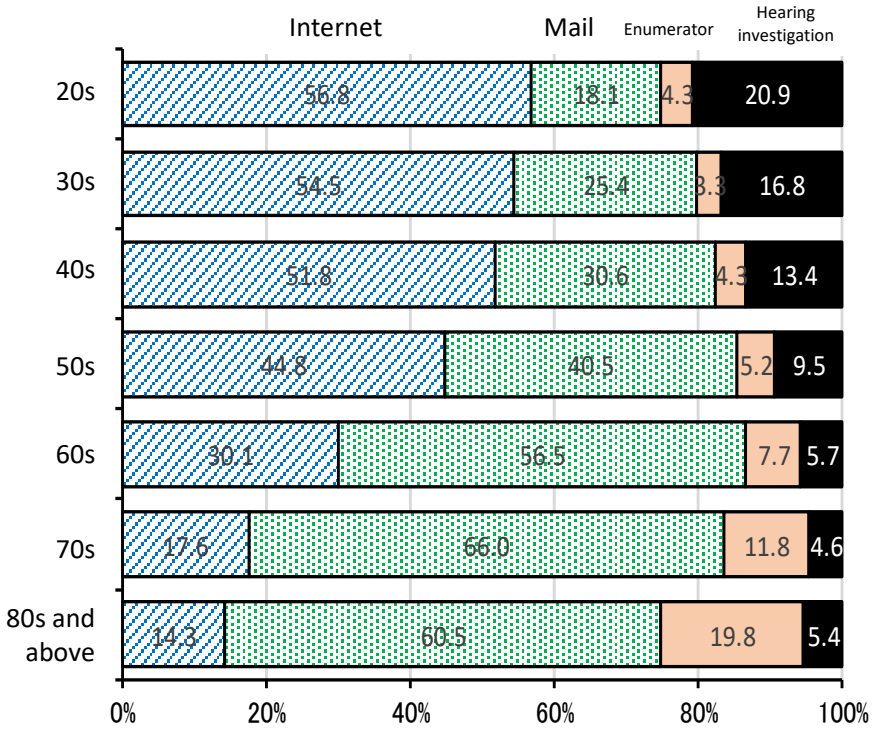
Household attribution was evaluated for each response method through pilot surveys.

→ Ratio of households to which hearing investigation was implemented was high among one-person households and young people, so it is possible that the attribution of unknown responses is biased.

Composition of response method by household members in the pilot survey



Composition of response method by the age of the head of household in the pilot survey



1. Background (3)

Further increase in the number of unknown values may lead to the underreporting of results by attribution, as well as to a bias in their composition.

→ There is a risk of not only decreasing the convenience of results, the excessive number of unknown values may also result in the bias of tabulated results.

In the 2015 Population Census, reference tables were prepared for certain results by imputed values.

In the 2020 Population Census, reference tables with imputed values were further expanded and detailed. They were prepared with an emphasis on bias correction.

The 2020 Population Census
Efforts in Providing the Results Resolving the Problem of Unknown Values

1. Background
2. **Method of the imputation**
3. Results of the imputation
4. Future development

2. Method of the imputation (1)

2015 Population Census: Unknown values for age and nationality are imputed by proportional distribution for the Basic Complete Tabulation on the Population and Households.

2020 Population Census: As a process prior to proportional distribution, age and nationality are imputed at the level of individual questionnaire sheet for those where reasonable estimation is possible, and focused on the correction of bias for the Basic Complete Tabulation on the Population and Households.
Then, the remaining unknown values for age, nationality, and marital status are imputed by proportional distribution.

Further, unknown values for basic items are imputed by proportional distribution also for other tabulation types.

2. Method of the imputation (2)

2020 Population Census: Method of the imputation for the Basic Complete Tabulation on the Population and Households

<Preprocessing (partial imputation)>
Impute the age and nationality (Japanese or foreigner) at the level of individual questionnaire sheet for data where all basic items are unknown for each household unit.

<u>Two-or-more-person household</u>
→ Age and nationality are imputed by using the Hot-deck method for small area, composition of household members by sex and type of building.

<u>Single-person household</u>
→ Age and nationality are imputed by using the Cold-deck method, based on administrative record information (foreign resident registration data) by small area and sex.
Then, of the remaining unknown age data for people who live in privately rented apartment houses or flats, data is imputed stochastically by municipality and sex, based on estimated population.

2. Method of the imputation (3)

2020 Population Census: Method of the imputation for the Basic Complete Tabulation on the Population and Households

<Proportional distribution>
Cross-tabulation tables by municipality based on the main classification shown below are prepared from the data where unknown values remain after the partial imputation.

Type of household, Sex, Age, Nationality (Japanese or foreigner), Marital status, Type of building
*** Unknown values exist for the underlined items**

In each cross-tabulation table, unknown values of items to be proportionally distributed (underlined) are imputed by proportional distribution according to the composition ratio of the data excluding unknown values for each cross-tabulation type.

→ Released result tables are reconfigured from the cross-tabulation tables after the imputation.

2. Method of the imputation (4)

2020 Population Census: Method of the imputation for other tabulation types

Unknown values for main items (underlined) are also imputed by proportional distribution for other tabulation types.

Tabulation type	Main classification items
Basic Complete Tabulation on Labour Force	Type of household, Sex, Age (five-year groups), Labour force status, Industry, Occupation, Employment status
Tabulation on Place of Work or Schooling	Type of household, Sex, Labour force status, Place of work or schooling
Tabulation on Internal Migration	Type of household, Sex, Age (five-year groups), Place of usual residence five years ago

* The preprocessing partial imputation is not implemented for tabulation types other than the Basic Complete Tabulation on the Population and Households.
* As for the tabulation types including the Basic Complete Tabulation on the Population and Households, the results of 2015 were also retroactively tabulated through the same method and simultaneously released in the form of reference tables.

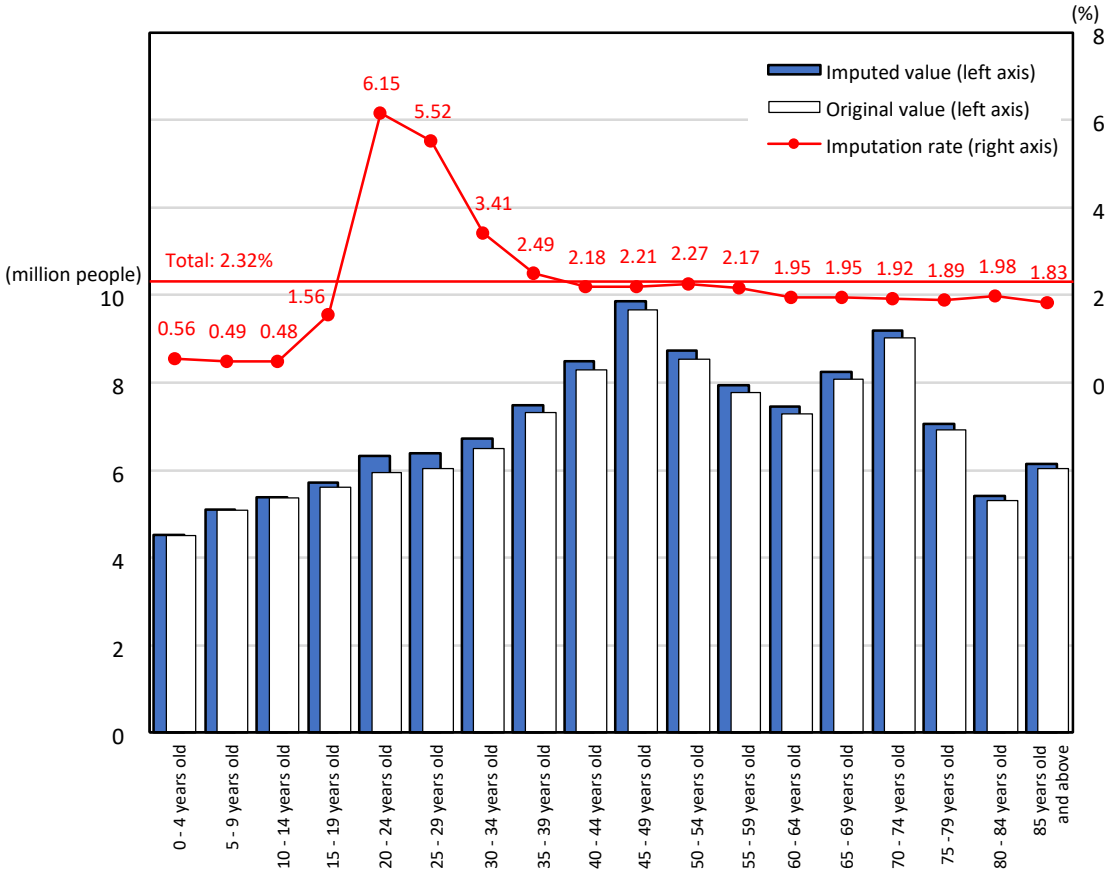
The 2020 Population Census
Efforts in Providing the Results Resolving the Problem of Unknown Values

1. Background
2. Method of the imputation
3. **Results of the imputation**
4. Future development

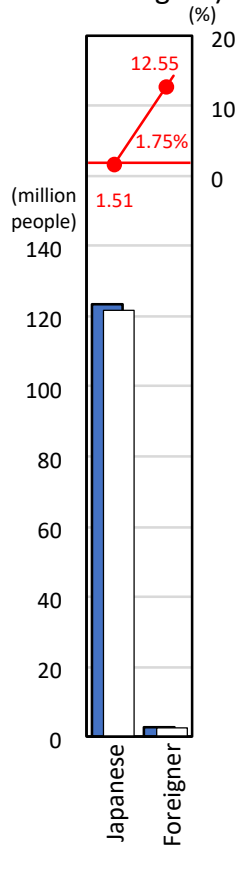
3. Results of the imputation (1)

Comparing the results before imputation (original value) and after imputation (imputed value), the results are intensively imputed among the young people, foreigners, and those never married, where it is assumed there are many targets that are difficult to survey.

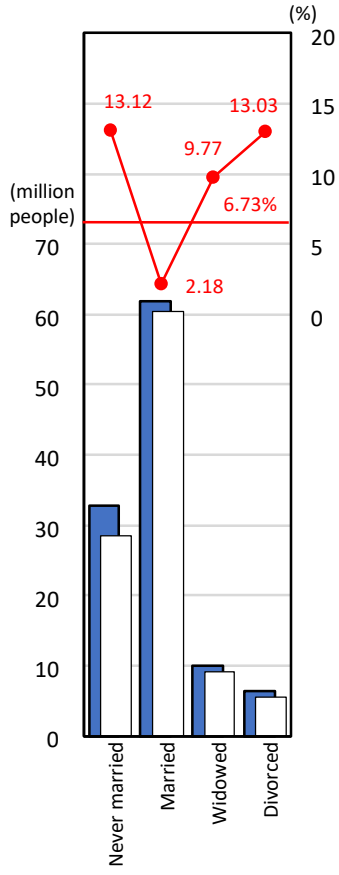
Population by age group (2020)



Nationality (Japanese or foreigner)

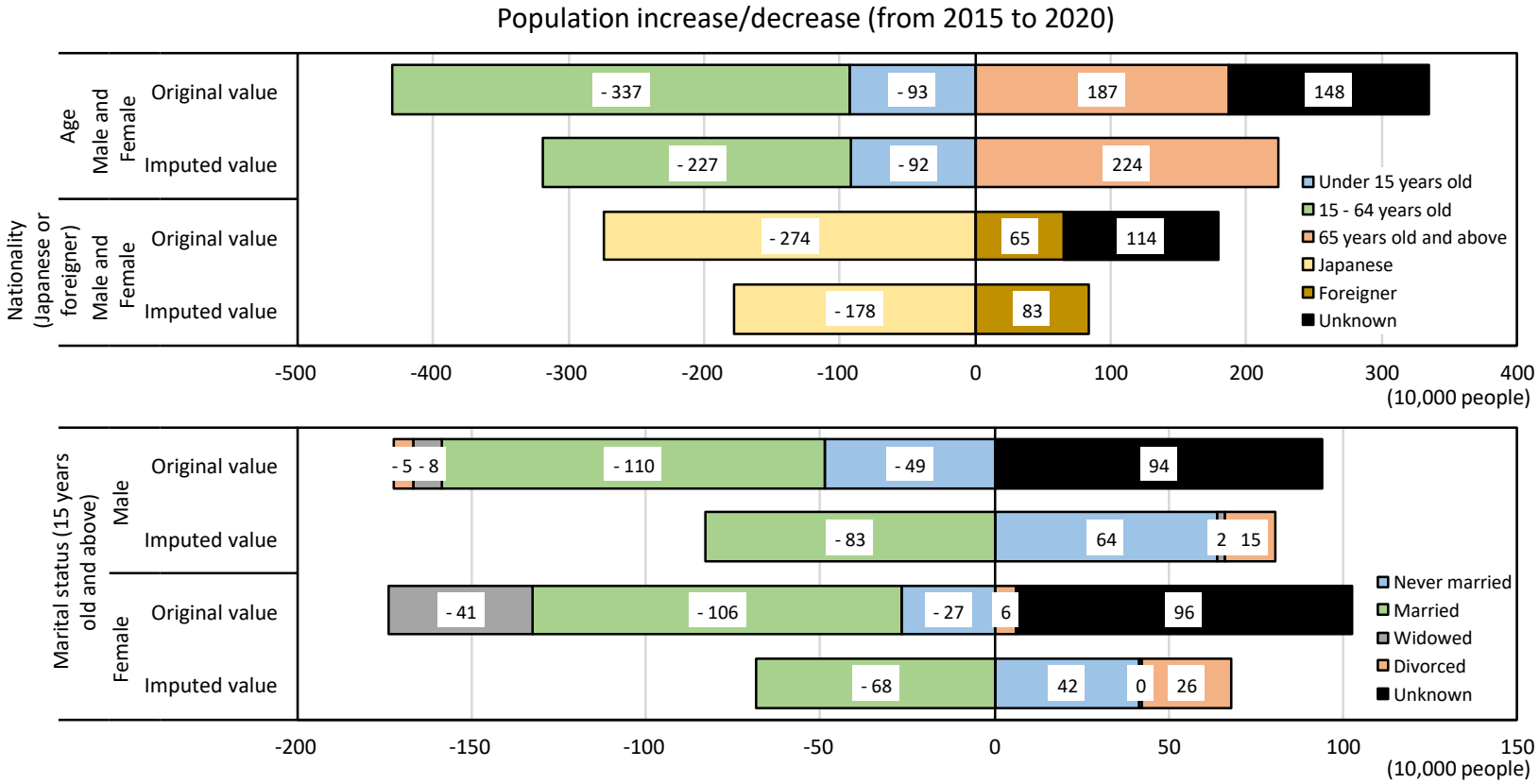


Marital status (15 years old and above)



3. Results of the imputation (2)

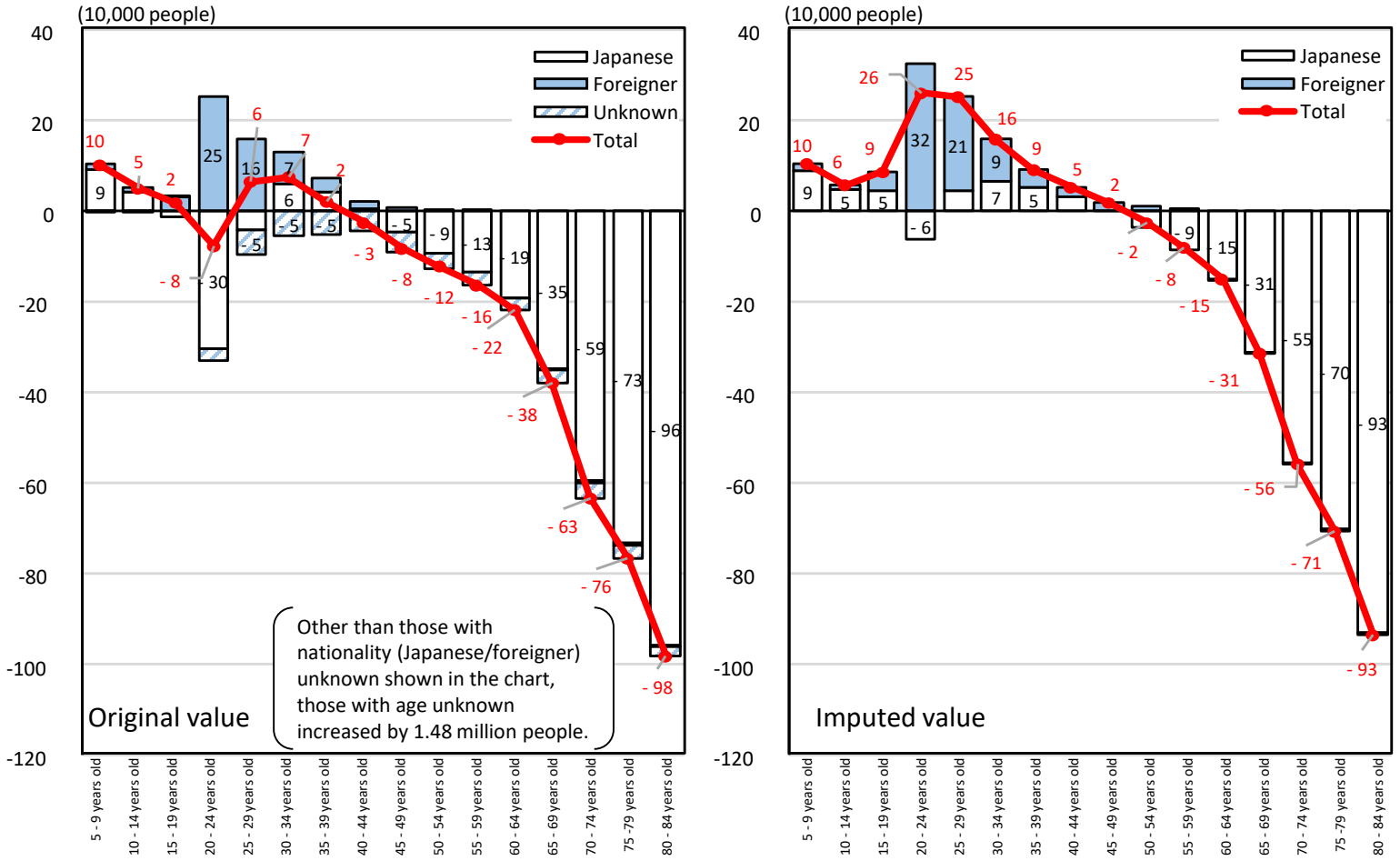
Looking at the increase/decrease of original values and imputed values from five years ago, as for imputed values, increase of unknown values is resolved, and decrease among each breakdown item was smaller while increase was larger than the original values.
In particular, the results of "Widowed" and "Never married" turned from decrease to increase.



3. Results of the imputation (3)

Increase/decrease from five years ago by age and nationality in terms of cohort shows that the decrease of young Japanese in the original value is resolved in imputed value.

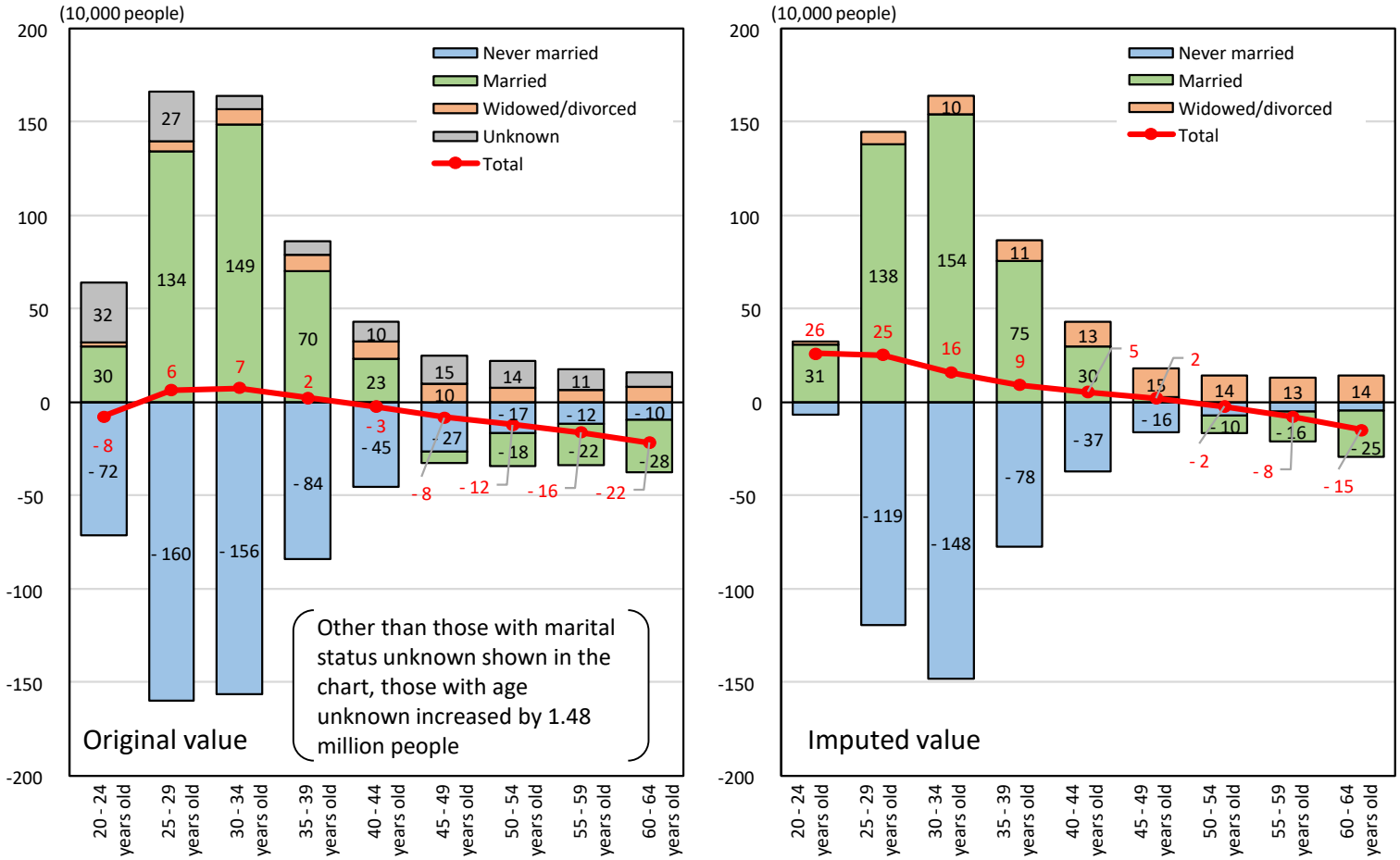
Cohort population increase/decrease (from 2015 to 2020)



3. Results of the imputation (4)

Increase/decrease from five years ago by age and marital status in terms of cohort shows that the decrease of young people “Never married” in the original value is resolved in imputed value.

Cohort population increase/decrease (from 2015 to 2020)



3. Results of the imputation (5)

The result with which unknown values are resolved and convenience increased will be actively utilized in materials transmitted to users by the Statistics Bureau, Ministry of Internal Affairs and Communications, including press release materials and analysis materials to users, as well as for the reference population used in other official statistics.

(Example) Population Estimates (Ministry of Internal Affairs and Communications)
Complete life tables (Ministry of Health, Labour and Welfare)

It is also scheduled to be used as the reference population in the Population Projection (National Institute of Population and Social Security Research) that will be estimated from now on.

The 2020 Population Census
Efforts in Providing the Results Resolving the Problem of Unknown Values

1. Background
2. Method of the imputation
3. Results of the imputation
4. **Future development**

4. Future development (1)

Because the imputed values for 2020 are basically calculated through proportional distribution, the following issues remain to be addressed.

- **Expansion of imputed items and tabulation result tables**
- **Dealing with Tabulation for Small Areas**
  → The precision of cross tabulation table used for proportional distribution has limitation in terms of realizing stable imputation.
- **Securing the consistency among different tabulation types**
  → Because proportional distribution is implemented independently for each tabulation type, consistency among different tabulation types cannot be secured.
- **Dealing with the secondary use of information on the questionnaire sheet**
  → Because the data is imputed on the cross-tabulation table, imputation result does not exist at the individual questionnaire sheet level, so secondary use is unavailable.

4. Future development (2)

As a potent means for solving these issues, it is assumed to impute all data at the individual questionnaire sheet level.

Information collection and analysis on the actual examples of imputation and imputation systems in other countries are implemented.
Also considering international trends, technical study will be promoted in the future as well.

Aim to further upgrade the imputation of unknown values and improve convenience for the 2025 Population Census.