DE GRUYTER
OPEN

Journal of Official Statistics, Vol. 31, No. 2, 2015, pp. 231–247, http://dx.doi.org/10.1515/JOS-2015-0015

Quality Assessment of Imputations in Administrative Data

Matthias Schnetzer[1], *Franz Astleithner*[2], *Predrag Cetkovic*[2], *Stefan Humer*[2], *Manuela Lenk*[3], *and Mathias Moser*[2]

This article contributes a framework for the quality assessment of imputations within a broader structure to evaluate the quality of register-based data. Four quality-related hyperdimensions examine the data processing from the raw-data level to the final statistics. Our focus lies on the quality assessment of different imputation steps and their influence on overall data quality. We suggest classification rates as a measure of accuracy of imputation and derive several computational approaches.

Key words: Administrative data; register–based statistics; quality assessment; imputations.

1. Introduction

Not only has the importance of administrative data as input for statistical purposes increased steadily over the last decades (Eurostat 2003), but the 2010 round of housing and population censuses has also provided an even stronger focus on register-based statistics on a European level. Following the Scandinavian countries, approximately one third of the United Nations Economic Commission for Europe (UNECE) members moved from a traditional to a (partly) register-based approach for their censuses (UNECE 2014). This transition was associated not only with a substantial change in the workflow of data generation, checking, and dissemination, but also with the need for new concepts to evaluate the quality of this kind of data.

Accordingly, a great number of books and articles concerning data quality issues has been published. Departing from Pipino et al. (2002), Batini and Scannapieco (2006), and Karr et al. (2006) who emphasize the role of a broad and multidimensional understanding of data quality in general, UNECE and EUROSTAT (2006), and Wallgren and Wallgren (2007) formulated guidelines for looking more closely on the dimensions of data quality in the National Statistical Institutes (NSI). In the following years, the NSIs developed different implementations of the CES recommendations; see Daas et al. (2008) for an early or UNECE (2014) for a recent review. Due to their 30 to 40 years of experience with register–based statistics, the approaches to quality assessment developed by the Nordic countries (UNECE 2007; Zhang 2011; Daas et al. 2012; Hendriks 2012; Axelson et al. 2012;

[1] Chamber of Labour Vienna – Department of Economics, Prinz-Eugen Str. 20–22, 1040, Vienna, Austria. Email: matthias.schnetzer@akwien.at
[2] Vienna University of Economics, Welthandelsplatz 1, 1020, Vienna, Austria. Emails: franz.astleithner@wu.ac.at, predrag.cetkovic@wu.ac.at, shumer@wu.ac.at and matmoser@wu.ac.at
[3] Statistics Austria, Unit Register-based census, Guglgasse 13, A-1110, Vienna, Austria. Email: manuela.lenk@statistik.gv.at

Zhang 2012) are a natural starting point for further developments. However, there is only scant research on the quality assessment of imputations within administrative data sources. In this article, we want to contribute to filling this research gap by introducing measures to evaluate imputation quality and embed this assessment into the broader quality framework put forward in Berka et al. (2010).

The remainder of this article proceeds as follows. First, we will introduce the general quality framework for the Austrian census in 2011 in Section 2. Section 3 proposes a structural approach for the quality assessment of imputations which is illustrated by examples from the Austrian census in Section 4. Finally, Section 5 offers some concluding remarks.

2. Basics of a Quality Framework for Administrative Data

2.1. The Principle of Redundancy in Register–Based Statistics

A major quality-related concern of register-based statistics is the selection of appropriate data sources as a supply for required information. In such a setting, the principle of redundancy ensures adequate data quality by collecting information from as many sources as possible. This can be achieved as multiple registers deliver autonomous information on a common subject and serve as instruments for cross-checks and validation.

Figure 1 illustrates the principle of redundancy with an example from the Austrian register-based census. Seven base registers are connected to the respective topics of the census. The three "backbones" of the census are the Central Population Register (CPR), the Housing Register of buildings and dwellings (HR) and the Business Register of enterprises including their local units (BR). All of these registers were set up after 2001 during the phase of transition from a traditional to a register-based census. These registers determine the total number of buildings and dwellings, the number of enterprises and the number of persons with their main residence in Austria. Some registers like the CPR, the HR or the BR are maintained by Statistics Austria. Others, like the Central Social Security Register (CSSR) or the Unemployment Register (UR) are kept by external data holders.

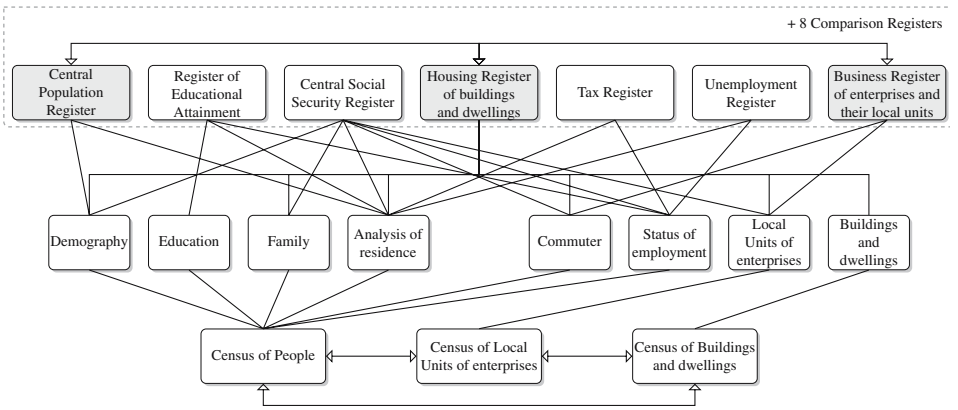

Fig. 1. The principle of redundancy in the Austrian register-based census.

To assure data quality, the base registers are backed up by eight comparison registers. The data for these comparison registers are provided by 35 external data holders and are mainly used for validation and cross checks. In this setup one specific base register is chosen to provide the basic information for a certain census variable, whereas the comparison registers are used to confirm these values (Lenk 2008). However, the comparison registers in some cases provide data that are not or only partly available in the base register. Due to the autonomous data administration of the external data holders, the sources might also deliver contradictory values for the same attribute (Berka et al. 2010; Kapteyn and Ypma 2007) or even use different concepts. Given potentially different values across the registers, one data source is chosen by a predefined ruleset in order to provide the required information. For this purpose, Statistics Austria maintains a ranking of registers that is ordered by confidence in the register quality based on long-time experience with the data. In the future, the results of the quality framework will offer the possibility to control and possibly adapt this ruleset.

Accordingly, this approach implements the principle of redundancy and ensures better quality by acquiring the information from as many registers as possible. However, despite a large number of different sources, missing values may still occur in the data which require special attention.

2.2. *The Quality Assessment of Administrative Data*

As far as this principle of redundancy is concerned, the NSI has little influence on the data maintenance at the external data holders who provide major parts of the required information. This emphasizes the quality assessment in the process of register–based statistics (Daas and Fonville 2007; Lanzieri 2009; Laitila et al. 2011). The definition of "quality" in this context includes multiple perspectives. According to the European Statistical System (ESS), the assessment of quality should consider *relevance*, *accuracy* and *reliability*, *timeliness* and *punctuality*, *accessibility* and *clarity* as well as *coherence* and *comparability*.

Herzog et al. (2007) provide a similar enumeration and add *completeness* as another quality dimension. Moreover, they list several articles with alternative lists of properties.

Based on these criteria, there have been several approaches to quality-related frameworks which satisfy all or at least parts of these requirements (Daas et al. 2009; Iwig et al. 2013). In this contribution we briefly sketch a framework for the analysis of administrative data which has been introduced earlier (Berka et al. 2010, 2012).

The quality framework for administrative statistics consists of three levels: the raw data (i.e., the base and comparison registers), the combined dataset (Central Database, CDB) and the imputed dataset (Final Data Pool, FDP). Figure 2 illustrates the process, which starts with the delivery of raw data from the various administrative data holders. If there are multiple sources that provide information on the same attribute, the data is connected via a unique personal key and merged to data cubes in the CDB (e.g., attribute A). Finally, the CDB is enriched with imputations for item nonresponse and implausible values. This step leads to the Final Data Pool, which now contains both real and estimated values. This standardized quality assessment enables the NSI as well as the data holders and end users to compare and monitor the quality a) between the different registers,

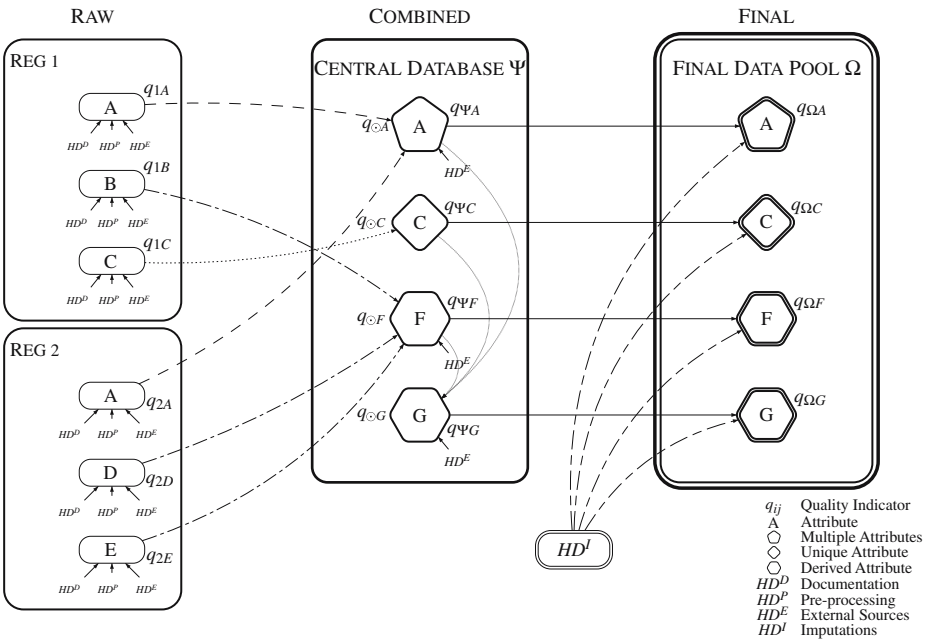

Fig. 2. Quality framework for register–based censuses.

b) throughout the data processing and c) between census generations for each attribute on the individual level.

In this framework, quality information at the raw data level (see left boxes in Figure 2) is obtained via three hyperdimensions: *Documentation* (HD^D), *Preprocessing* (HD^P) and an *External Source* (HD^E). These three measures are standardized between zero and one, where a higher value implies better data quality. These indicators are calculated for each attribute in each register.

The first hyperdimension HD^D describes quality-related processes at the register authority as well as the documentation of the data (metadata) prior to seeing the data. To achieve this, an estimate of confidence and reliability for each data holder is generated through the use of a questionnaire containing nine scored questions. These questions cover four main topics: data historization, definitions, administrative purpose, and data treatment. As an example, the register authorities are asked to evaluate the timeliness of their data on an ordinal scale. The compatibility of definitions, however, is evaluated as either given or not given on a dichotomous scale. Based on the nine questions, a quality measure is computed as the obtained score divided by the maximum achievable score. As a result, the NSI is able to check for data collection methods or legal enforcements of data recording which may significantly influence the data quality. A detailed description of the questions and the calculation of the hyperdimension HD^D is given in Berka et al. (2010).

The second hyperdimension HD^P is concerned with formal errors in the raw data. This covers range errors, item nonresponse, and missing primary keys, which are detected in this step of the quality framework. The final result of this hyperdimension is given by the ratio of *usable records* to the *total number of records*. Again, this procedure is carried out for each attribute in each register. At this stage plausibility checks for inconsistent

combinations of values are not considered, since this measure only covers information on the formal correctness of values of a single attribute in a single register. If the proportion of usable records for an attribute in a certain register is smaller than that of the same attribute within another register, the respective quality measure will accordingly be lower. The unusable records in the registers are considered as missing data in the following.

Finally, the third hyperdimension HD^E provides a comparison between the register–based data and an independent external source. In Austria, the microcensus is a common benchmark for representative surveys and is the best comparative dataset available. Thus a check for consistency with this external source results in the third quality measure, which is the ratio of the *number of consistent values* to the *total number of linked records* between the register and the external source.

Given these three quality measures, an overall quality indicator for each attribute in each register can be derived from these hyperdimensions as a weighted average. Hence, the quality measure does not differ between the different statistical units of the same register. In our framework, each hyperdimension is believed to have the same weight and therefore an equal impact on quality. For other applications these weights may be chosen differently, that is, according to their assigned importance. The resulting value summarizes the existing quality-related information for each attribute in each register. Hence, this indicator is able to capture quality-related effects throughout the process of data generation through to the raw data available in the registers.

The registers are then combined in the Central Database (CDB, center box in Figure 2) which covers all attributes of interest for the register–based statistics. At this level a quality indicator for each attribute for each statistical unit is computed for the first time. Concerning the quality evaluation for the CDB, we distinguish three types of attributes by their origin:

- *Unique attributes* exist in exactly one register, for example educational attainment (see attribute C). Accordingly, the quality measure in the CDB will be the same as in the raw-data.
- *Multiple attributes* show up in several registers, for example sex (see attribute A). The information from multiple sources has to be combined. A predefined ruleset picks the most appropriate value for the CDB according to the constellation of the values in the raw-data registers.
- *Derived attributes* are created based on different attributes for example, current activity status (see attributes F and G). The registers do not contain any information for these attributes in the required specification but there is related information that can be used to derive the attribute of interest.

For a unique attribute, the quality indicator is the same across all statistical units equal to the raw-data quality measure, since only one source is available and therefore no conflicting or confirming evidence can arise. However, for multiple and derived attributes the values for each statistical unit may differ according to the quality values in the different origin registers. For missing values the quality indicator of an attribute is set to zero for the corresponding record, irrespective of the type of the attribute. A detailed description of the quality assessment for the three types of attributes in the CDB can be found in Berka et al. (2010, 2012).

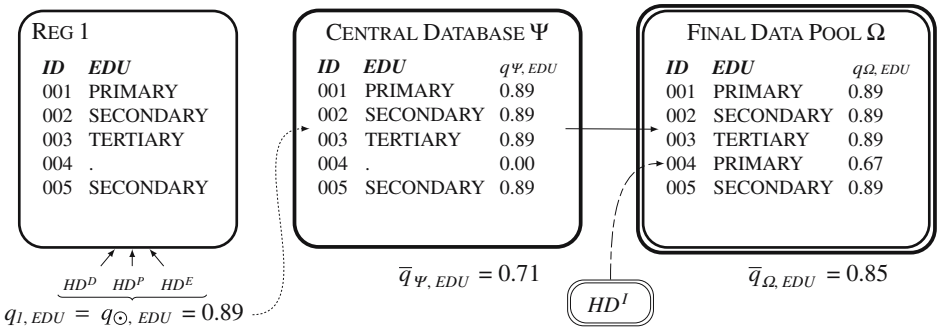

Fig. 3. Quality assessment for a unique attribute.

2.3. *Quality Assessment from the Central Database to the Final Data Pool*

In the previous section, we have examined the quality assessment on the raw-data level as well as in the Central Database. So far, item nonresponse has not been considered. However, missing and implausible values should be replaced by plausible information in the Final Data Pool (FDP, right box in Figure 2).

Figure 3 illustrates an artificial example of the quality assessment for a unique attribute, educational attainment (*EDU*) with a missing value in the raw data, indicated by a dot. The quality measurement for this attribute starts with the application of the three hyperdimensions on the register level (HD^D, HD^P, HD^E). In this simple case, we consider a unique attribute so that the information from the register is directly transferred to the CDB because there is no other source to be considered. As a consequence, the quality indicators for the values on the CDB level remain unchanged. In our hypothetical example, $q_{\odot}$ equals 0.89, which is the result of a weighted average of the three hyperdimensions at the register level. If a record has no information or implausible information for an attribute, the quality indicator is set to zero on CDB level (see ID 004 in the center of Figure 3), since we do not have any valid information for this record. However, the quality for this entry is updated after the imputation process. The corresponding quality measure in the FDP – we show its derivation in the following – is supposed to be 0.67 (HD^I). The entries which are not imputed have the same quality in the FDP as in the CDB. The average of the quality measures for all records on FDP level results in the final quality indicator of $q_{\Omega} = 0.85$ for educational attainment. Thus the average of the quality indicator for this attribute after the imputation procedure $\bar{q}_{\Omega,EDU}$ increases as compared to the average quality on CDB level $\bar{q}_{\Psi,EDU}$.

This procedure is applied not only for unique attributes but for all three types of attributes. Moreover, it leads to the same result: since imputations provide plausible values for attributes that would otherwise deliver no information for a certain record, the overall quality of the attribute increases after the application of the imputation process. The magnitude of this increase depends however on the quality of the imputation process as measured by HD^I. In the following, we will focus on how the hyperdimension Imputation (HD^I) can be derived.

3. A Structural Approach to the Quality Assessment of Imputations

3.1. Imputation Process and Estimating Order

Due to the principle of redundancy, the amount of missing values in register-based statistics is generally considered to be quite low, since a large number of variables are covered in multiple registers. For instance, in the Austrian register-based census of 2011, the level of item nonresponse for most attributes does not greatly exceed ten percent. For demographic variables, like sex or age, the number of missing values is considerably lower. Nevertheless, some values need to be imputed due to a variety of reasons.

EU Commission Regulation 1151/2010 distinguishes between *item imputation* and *record editing* (European Commission 2010). Item imputation refers to the insertion of artificial but plausible information into a data record with a missing value in this specific attribute. More specifically, imputations try to set a value in accordance with information already available either in the same record or in the rest of the database. Record editing is the process of checking and modifying data records to tender them plausible while preserving major parts of these records. However, record editing is often accomplished by deleting implausible (or out-of-range) values and subsequently reimputing the missing entries. Conversely, Chambers (2001, 11) does not distinguish "between imputation due to missingness or imputation as a method for correcting for edit failure." He argues that in both cases the true values are missing. For quality assessment in Austria, both types are treated the same way irrespective of the reason for the imputation.

For example, Chambers (2001, 11f) distinguishes between five quality-related properties that imputations should fulfil:

1. Predictive Accuracy: the imputed values should be as "close" as possible to the true values.
2. Ranking Accuracy: the imputation process should preserve the order of imputed values (for attributes which are at least ordinal).
3. Distributional Accuracy: the imputation procedure should preserve the distribution of the true data values.
4. Estimation Accuracy: the lower order moments of the distribution of the true values should be reproduced by the imputation process (for scalar attributes).
5. Imputation Plausibility: the imputation procedure should result in imputed values that are plausible.

These conditions may serve as a reference point for the quality assessment of imputations. Furthermore, the imputation procedure requires a hierarchical estimation order to connect all necessary steps in a chronological way. In this respect, two aspects have to be considered on a theoretical basis (Kausl 2012):

- In most statistics based on administrative data, a variety of registers is used in order to ensure sufficient quality for all required attributes. Due to possible differences in the data delivery (delays) it is necessary to check at which time each item can be edited.
- The choice of predictors used for imputations should be based on their association with the variables to be imputed. Therefore, it is imperative to analyze the highest correlations between the variables to develop optimal estimation models for each

imputation step. Variables that have already been imputed can be used as predictors to estimate other items.

As an example, Figure 4 illustrates the imputation interdependencies between the variables of the Austrian census topics. The hierarchical work flow is indicated by the arrows from one attribute to another. The relationships between variables are not confined within the topics (e.g., *LMS* ← *AGE, SEX* and *POB*), but also connect variables between the topics (e.g., *EDU* ← *AGE, SEX, COC* and *PFE*). Demographic attributes, like age and sex, are the first ones in the estimation order, variables concerning the labour market are the last. Accordingly, many other variables are required to impute missing values in labour market variables, such as occupation (*OCC*). In the next step, the quality of the imputations has to be evaluated.

3.2. *Applied Imputation Methods*

For the quality assessment of imputations in register–based statistics (HD^I in our framework), the distinction of methods is crucial (see Kausl 2012). We distinguish

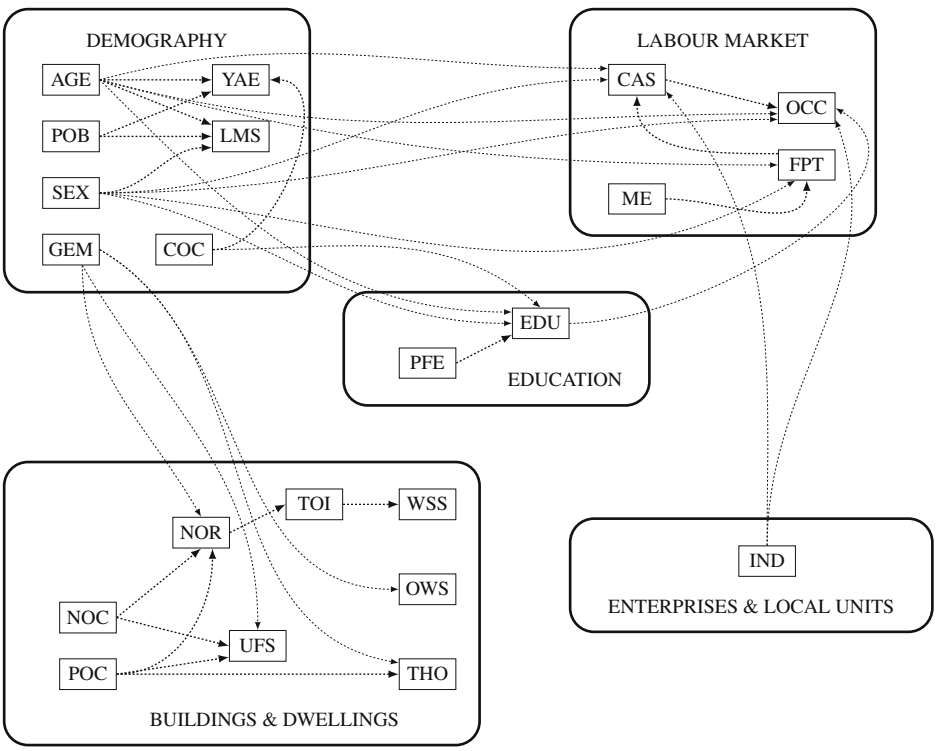

Fig. 4. Structured work flow of the imputation process. POB = place of birth, LOC = size of locality, COC = country of citizenship, YAE = year of arrival in the country since 1980, LMS = legal marital status, CAS = current activity status, ME = marginal employment, FPT = full-part-time employment, OCC = occupation, PFE = participation in formal education, EDU = educational attainment (highest completed level), NOC = number of occupants, POC = period of construction, NOR = number of rooms, TOI = toilet/bath facilities, WSS = water supply system, UFS = useful floor space, OWS = type of ownership, THO = type of heating, IND = industry (branch of economic activity).

between *deterministic editing*, *statistical estimation* (primarily hot-deck technique, but also logistic regressions) and *statistical matching*. To begin with the first, missing values in the data can be imputed by deterministic rules, even before applying statistical methods, because we are able to derive missing values from auxiliary data. Two examples from the Austrian register-based census illustrate such cases:

- Missing values in the legal marital status (*LMS*) are classified according to the Central Social Security Register. Information on individuals receiving a widow's or widower's pension is provided by this register. The relevant information is gathered to change missing values of the attribute *LMS* to *"widowed"* if a person receives a widow's or widower's pension.
- People younger than 15 years are classified *"not applicable (persons under 15 years of age)"* with regard to the educational attainment (*EDU*). Their current activity status (*CAS*) is *"persons below the age of 15"* and their marital status (*LMS*) is *"never married"*.

We do not consider such derivations with the utmost matching probability an estimation in the narrower sense, but rather see them as plausibility steps. However, there are also derivations with substantial uncertainty due to a lack of information. Nevertheless, in the following cases taken from the Austrian register-based census no statistical imputation method is necessary:

- The Central Population Register has information on the place of birth (*POB*). Missing values are filled with information on the country of citizenship (*COC*) if the person has foreign citizenship. The available data justify this assumption: 77 percent of individuals with a foreign *COC* were also born in this (foreign) country. Hence, even though there is uncertainty, this imputation method classifies 77 percent of the attribute *POB* as correct when it is applied to observed data for 2011.
- Suppose the marital status (*LMS*) is missing and there is another individual living in the same household. If the other person is *"married"*, the age difference between the two individuals is less than 18 years, and their sex differs, then the missing marital status is set to *"married"*.

Another important imputation method for the Austrian census is hot-deck imputation. This method choses the imputed value from an assumed or estimated distribution that is taken from existing data (Little and Rubin 2002). It is suitable for all scenarios of missing data, except for missing not at random higher than ten percent (Roth 1994). A detailed review of hot-deck methods is given by Andridge and Little (2010). In the Austrian case, individuals are aggregated to groups (*"decks"*) by attributes which are strongly correlated to the response variable. The distribution in the decks of the source data, derived from the FDP, is transferred to the corresponding group of the target data. Table 1 gives an example of artificial data for the *LMS*. The distribution of the existing values in the census of the same year is applied on the missing values for the same attribute. As an example, 55.6 percent of all females aged 30 to 40 years with their main residence in the federal state Tyrol and a missing value for *LMS* will be considered married women. Since we cannot be sure which women with a missing LMS are actually married, a uniformly distributed random variable with the interval [0, 1] determines the assignment of the LMS. According to our example in Table 1, the interval [0, 0.37) is assigned to "LMS = never married",

Table 1. Artificial example of the deck for legal marital status (LMS)

Sex	Age	Federal state	Size of deck	$P_{never\ married}$ %	$P_{married}$ %	$P_{divorced}$ %	$P_{widowed}$ %
Female	30–40	Tyrol	50.000	37	55.6	7	0.4
Male	50–60	Vienna	100.000	12	66	20	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	

the interval [0.37, 0.926) is assigned to "married", the interval [0.926, 0.996) is assigned to "divorced" and finally the interval [0.996, 1) is assigned to "widowed".

Finally, statistical matching is the last applied imputation method in the Austrian census. It is based on the combination of two incomplete records. We will explain the procedure using the example of a missing observation for educational attainment (*EDU*). Register–based statistics rely on unique identification keys for every individual in order to combine the information from multiple data sources. Consider a data record with a missing value for educational attainment (*EDU*). Consider another data record with a missing unique identification key but information on several other attributes, including *EDU*. Statistical matching searches these loose observations and connects them with individuals who have a missing value for *EDU* but otherwise have the same characteristics. Two incomplete records, one of them useless because of the missing identification key, can be merged to one complete record.

3.3. *Quality Assessment of Imputation Models*

In general, an overall quality measure for imputations requires the evaluation of two parts, the input of the estimation model as well as the output (i.e., the accuracy of the model). The inputs of the estimation models are assessed with the three hyperdimensions HD^D, HD^P, and HD^E that are combined in the CDB.

For the evaluation of the model itself, the so-called *classification rate* Φ is used to obtain a quality measure for the imputations (Hui and AlDarmaki (2012) as well as Chambers (2001)).

It is a general measure for the goodness of fit and can also be calculated for a variety of imputation techniques. Its principle is to apply the imputation model to already existing data and compare the results of the imputation process with the true values of these observations. The classification rate equals the ratio between the matching values and the number of all compared entries.

This measure can be applied specifically to categorical variables and is shown in equation (1), where $\hat{Y}_i$ is the estimated value for the observed value Y_i^* of person i, n is the sample size and I is an indicator function. Take the legal marital status (*LMS*) as an example for a categorical variable. In this case the quality assessment should measure the hit ratio, that is, the probability that the estimation model picks exactly the right category of the true value.

$$\Phi = 1 - n^{-1} \sum_{i=1}^{n} I(\hat{Y}_i \neq Y_i^*) \tag{1}$$

For ordinal variables, the distance of the imputed value to the true value is relevant, hence Equation (2) is a modification of the classification rate Φ that measures and standardizes this gap. A satisfactory quality indicator has to consider the accuracy of the model, which means measuring the contiguity of the estimated value to the true value. Assume several categories of the attribute educational attainment (*EDU*), ranging from primary to higher tertiary education. If the true value was *higher tertiary education*, an estimated value of *lower tertiary education* would be more accurate than an estimated value of *lower secondary education*.

$$\Phi = 1 - n^{-1}\sum_{i=1}^{n}\left(\frac{1}{2}\left[\frac{|\hat{Y}_i - Y_i^*|}{max(Y) - min(Y)} + I(\hat{Y}_i \neq Y_i^*)\right]\right) \tag{2}$$

For the case of numerical variables both concepts (1) and (2) can be applied. Chambers (2001, 15) suggests that the methods which are developed for categorical variables could also be applied to scalar attributes by first categorizing them. If the arbitrariness of categorizing variables should be avoided, an applicable imputation performance measure has to be constructed. However, a simple correlation coefficient between estimated and true values is considered to be a rather intuitive approach. One example for a metric attribute is the variable "useful floor space" (*UFS*) of a household. The correlation coefficient between the estimated and the true *UFS* can be applied analogously to the classification rate for the evaluation of the imputation model.

Finally, we provide an application of the major steps for the quality assessment of imputation methods described above: deterministic editing with and without uncertainty, statistical estimation as well as statistical matching. As already mentioned, the source variables for the imputation process are the attributes in the FDP rather than attributes in the raw data. Therefore, the quality indicator from the FDP delivers the quality information for the source variables, with the values being used for the single statistical units. According to the type of imputation we distinguish between the following quality assessment rules:

- **Deterministic editing without uncertainty:** the input quality equals the quality of the source variables $q_{\Omega,i}$ where i denotes the attribute. The output quality equals 1, as there is no uncertainty about the correctness of the model. The overall quality of the imputation yields

$$HD^I = \underbrace{\frac{1}{n}\sum_{i=1}^{n} q_{\Omega,i}\cdot\Phi}_{\overline{q}_{Input}} \tag{3}$$

  where $\Phi = 1$.

- **Deterministic editing with uncertainty:** the input quality again equals the average quality of the source variables $q_{\Omega,i}$, while the output quality equals the classification rate Φ, as shown in Equation (3).
- **Statistical estimation:** we define imputation quality as the average quality of the predictors $q_{\Omega,i}$ (input quality) times the classification rate (output quality) for the imputations (again see, Equation 3). This measure is independent of the number of

Table 2. Results for the calculation of HD^I for legal marital status (LMS)

	Deterministic editing*				Hot decks											
	1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{q}_{\Omega,CAS}$	0.914															
$\bar{q}_{\Omega,AGE}$		0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
$\bar{q}_{\Omega,SEX}$			0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
$\bar{q}_{\Omega,POB}$						0.983		0.983		0.983	0.983	0.983	0.983		0.983	0.983
$\bar{q}_{\Omega,HH-ID}$											0.821					
$\bar{q}_{Input}$	0.914	0.998	0.999	0.999	0.999	0.994	0.999	0.994	0.999	0.999	0.951	0.994	0.994	0.999	0.994	0.994
Φ	1	1	1	1	0.928	0.679	0.674	0.894	0.891	0.783	0.84	0.838	0.77	0.734	0.716	0.709
HD^I	0.914	0.998	0.999	0.999	0.927	0.675	0.673	0.888	0.889	0.778	0.798	0.833	0.765	0.733	0.711	0.704

Source: Austrian register–based census 2011.

Note: *CAS* = current activity status, *POB* = place of birth, *HH_ID* = Household ID.

*Deterministic Editing: For example, if age is under 15 years then *LMS* is *"never married"*.

predictors and includes both the quality of the data used for the imputations as well as their ex-post fit.

- **Statistical matching:** two incomplete records – one without unique identification key, another one with the missing value – are merged. Therefore, no imputation in the narrower sense is carried out. The formerly missing value in the merged records from now is treated the same as any other non-missing value. The quality measure is obtained via the quality of the used data source.

4. Results of the Quality Measurement of Imputations in the Austrian Census

Applying the suggested quality assessment for imputations, we provide results for the attribute legal marital status (*LMS*) from the Austrian register–based census of 2011. Table 2 shows results for the used imputation methods, deterministic editing without uncertainty and hot-deck techniques.

Hot-deck imputations for the *LMS* are based on four variables. These predictors are age, sex, place of birth (*POB*) and household identification number (*HH_ID*). For every deck the average input quality $\overline{q}_{Input}$ of the source variables is computed (on the level of statistical units). Table 2 shows the twelve imputation decks in the columns. The classification rate Φ measures the hit ratio when applying the imputation method on observed data to evaluate the model accuracy. Thus we obtain a classification rate for every deck. The quality of the source variables is multiplied with the classification rate, which results in a quality measure HD^I for the imputations. Table 2 presents the average of this measure for each deck.

Deterministic editing without uncertainty also uses other variables to gain information for the missing value. For example, the *LMS* can be derived from the variable age, as is shown in the second column of Table 2. The conditions stated by deterministic editing without uncertainty are true by definition. Therefore, HD^I equals the quality of the source variable(s) since the hit ratio is 1. Again, every statistical unit receives a quality measure for the input values and for the resulting HD^I.

The overall quality of the attribute *LMS* is positively affected, as can be seen in Table 3. The quality indicator $\bar{q}_{\Psi,LMS}$ in the CDB equals 0.955. Remember that the quality of missing values is set to zero in the CDB. In the FDP, imputed entries receive the quality of HD^I which leads to an increase of the overall quality $\bar{q}_{\Omega,LMS}$.

5. Conclusion and Outlook

This article provides a structural approach for the quality assessment of imputations. The evaluation of several imputation methods is embedded in a broader quality framework for register-based statistics as applied by Statistics Austria. Thus, analyzing the imputation process is one component in achieving a comprehensive and conclusive quality assessment for statistics based on administrative data. While other quality dimensions, such as metadata or formal errors, have received a lot of attention in the literature, there is less international experience on the evaluation of imputation procedures.

The quality assessment of imputations is the last conceptual step in the general framework introduced by Berka et al. (2010) and extended in Berka et al. (2012) for the register-based census in Austria. In this contribution, we address the possibility of

Table 3. Quality assessment from CDB to FDP for legal marital status (LMS)

		Deterministic editing				Hot decks												
	CDB	1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12	FDP
n	7,700,274	22925		457918		72091	56115	60	25540	79	3019	56269	547	124	73	6852	54	8,401,940
$\bar{q}_{\psi,LMS}$	0.955																	
HD^1		0.914	0.998	0.999	0.999	0.927	0.675	0.673	0.888	0.889	0.778	0.798	0.833	0.765	0.733	0.711	0.704	
$\bar{q}_{\Omega,LMS}$																		0.958

Source: Austrian register–based census 2011.

different scale levels of the imputed attributes which require different quality measurement approaches. The application of classification rates enables us to assess the input variables as well as the model accuracy of various imputation methods.

The framework suggests that the quality of an attribute increases after the imputation process, since the quality of a missing value equals zero. Thus the imputed values have a slightly positive effect on the overall quality indicators, as preliminary results for the Austrian census 2011 confirm. However, the impact of imputations on the quality measures is moderate, since the number of missing values is rather low in the vast majority of attributes in the census.

We are confident that this approach has various practical applications, as the derivation of overall quality measures is of use not only for the NSIs but also for the data holders and end users. In particular, the end users of register data are able to track the quality of particular datasets and can assess the reliability of certain attributes. Moreover, this framework may help NSIs to meet the EUROSTAT quality reporting requirements for register-based censuses. In the future, the results provided by the quality framework may influence the choice of the data source in case of contradicting values between contributing registers.

Given the increasing relevance of register–based statistics in the past decade, our outlook clearly indicates that the importance of administrative data will rise in the coming years. At the same time, quality management plays an essential role in order to justify the replacement of traditional statistics and surveys. The general framework presented in this article facilitates the monitoring of quality assessment over time; however, there are still tasks ongoing and further research ahead, such as the detailed quality measurement of statistical matching.

6. References

Andridge, R.R. and R.J.A. Little. 2010. "A Review of Hot Deck Imputation for Survey Non-response." *International Statistical Review* 78: 40–64. Doi: http://dx.doi.org/10.1111/j.1751-5823.2010.00103.x.

Axelson, M., A. Holmberg, I. Jansson, P. Werner, and S. Westling. 2012. "Doing a Register-Based Census for the First Time: The Swedish Experiences." In Proceedings of JSM 2012, Survey Statistics Section, July 28-August 2, 2012. 1481–1489. Alexandria, VA: American Statistical Association.

Batini, C. and M. Scannapieco. 2006. *Data Quality: Concepts, Methodologies and Techniques*. New York: Springer.

Berka, C., S. Humer, M. Lenk, M. Moser, H. Rechta, and E. Schwerer. 2010. "A Quality Framework for Statistics based on Administrative Data Sources using the Example of the Austrian Census 2011." *Austrian Journal of Statistics* 39: 299–308.

Berka, C., S. Humer, M. Lenk, M. Moser, H. Rechta, and E. Schwerer. 2012. "Combination of Evidence from Multiple Administrative Data Sources: Quality Assessment of the Austrian Register-Based Census 2011." *Statistica Neerlandica* 66: 18–33. Doi: http://dx.doi.org/10.1111/j.1467-9574.2011.00506.x.

Chambers, R. 2001. "Evaluation Criteria for Statistical Editing and Imputation." *National Statistics Methodological Series* 28: 1–41.

Daas, P., J. Arends-Tóth, B. Schouten, and L. Kuijvenhoven. 2008. "Quality Framework for the Evaluation of Administrative Data." In Proceedings of Q2008 European Conference on Quality in Official Statistics. 8–11 July, 2008. p 1–10. Available at: http://q2008.istat.it/sessions/paper/21Daas.pdf (accessed May 6, 2015).

Daas, P. and T. Fonville. 2007. *Quality Control of Dutch Administrative Registers: An Inventory of Quality Aspects*. Presented at the seminar on Registers in Statistics-Methodology and Quality, 21–23 May, 2007, Helsinki. Available at: http://www.oecd.org/std/38883259.pdf (accessed May 6, 2015).

Daas, P., S. Ossen, R. Vis-Visschers, and J. Arends-Tóth. 2009. "Checklist for the Quality Evaluation of Administrative Data Sources." Discussion Paper No. 09042, Statistics Netherlands. Available at: http://www.cbs.nl/nr/rdonlyres/0dbc2574-cdae-4a6d-a68a-88458cf05fb2/0/200942x10pub.pdf (accessed May 6, 2015).

Daas, P.J., S.J. Ossen, M. Tennekes, and E.S. Nordholt. 2012. "Evaluation of the Quality of Administrative Data used in the Dutch Virtual Census." In Proceedings of JSM 2012, Survey Statistics Section, July 28-August 2, 2012. 1462–1472. Alexandria, VA: American Statistical Association.

European Commission. 2010. "Commission Regulation (EU) No 1151/2010." *Official Journal of the European Union* 53 L 324: 1–10. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32010R1151&from=EN (accessed May 6, 2015).

Eurostat. 2003. "Quality Assessment of Administrative Data for Statistical Purposes." Contribution for the Working Group "Assessment of Quality in Statistics", October 2–3, 2003. Luxembourg. Available at: http://unstats.un.org/unsd/EconStatKB/Attachment323.aspx (accessed May 6, 2015).

Hendriks, C. 2012. "Input Data Quality in Register Based Statistics–the Norwegian Experience." In Proceedings of JSM 2012, Survey Statistics Section, July 28-August 2, 2012. 1473–1480. Alexandria, VA: American Statistical Association.

Herzog, T.N., F.J. Scheuren, and W.E. Winkler. 2007. *Data Quality and Record Linkage Techniques*. New York: Springer.

Hui, G. and H.I. AlDarmaki. 2012. "Editing and Imputation of the 2011 Abu Dhabi Census". Paper presented at the UNECE Work Session on Statistical Data Editing, Oslo, September 24–26, 2012. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2012/44_Abu_Dhabi.pdf (accessed May 6, 2015).

Iwig, W., M. Berning, P. Marck, and M. Prell. 2013. *Data Quality Assessment Tool for Administrative Data*. Technical Report, Office of Survey Methods Research (OSMR). Available at: http://www.bls.gov/osmr/datatool.pdf (accessed April 30, 2015).

Kapteyn, A. and J. Ypma. 2007. "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data." *Journal of Labor Economics* 255: 513–551.

Karr, A., A. Sanil, and D. Banks. 2006. "Data Quality: A Statistical Perspective." *Statistical Methodology* 3: 137–173. Doi: http://dx.doi.org/10.1016/j.stamet.2005.08.005.

Kausl, A. 2012. "The Data Imputation Process of the Austrian Register-Based Census." Paper presented at the UNECE Work Session on Statistical Data Editing, Oslo, September 24–26, 2012. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2012/41_Austria_Kausl.pdf (accessed May 6, 2015).

Laitila, T., A. Wallgren, and B. Wallgren. 2011. *Quality Assessment of Administrative Data*. Research and Development–Methodology Reports from Statistics Sweden, 2011:2, Statistics Sweden. Available at: http://www.scb.se/statistik/_publikationer/OV9999_2011A01_BR_X103BR1102.pdf (accessed May 6, 2015).

Lanzieri, G. 2009. *A Quantitative Approach to the European Union Census Quality Reporting*. Note by the Statistical Office of the European Communities. Available at: http://www.academia.edu/2565773/A_quantitative_approach_to_the_European_Union_census_quality_reporting (accessed April 30, 2015).

Lenk, M. 2008. *Methods of Register-based Census in Austria*. Technical Report, Statistik Austria, Vienna. Available at: http://www.statistik.at/web_de/static/methods_of_register-based_census_in_austria_055160.pdf (accessed May 6, 2015).

Little, R.J.A. and D.B. Rubin. 2002. *Statistical analysis with missing data*. New York: Wiley.

Pipino, L.L., Y.W. Lee, and R.Y. Wang. 2002. "Data Quality Assessment." *Communications of the ACM* 45: 211–218.

Roth, P.L. 1994. "Missing data: A Conceptual Review for Applied Psychologists." *Personnel Psychology* 47: 537–560. Doi: http://dx.doi.org/10.1111/j.1744-6570.1994.tb01736.x.

UNECE. 2007. *Register Based Statistics in the Nordic Countries. Review on Best Practices with Focus on Population and Social Statistics*. Technical Report, United Nations Economic Commission for Europe. Available at: http://unstats.un.org/unsd/dnss/docViewer.aspx?docID=2764 (accessed May 6, 2015).

UNECE. 2014. *Measuring Population and Housing–Practices of UNECE Countries in the 2010 Round of Censuses*. Technical Report, United Nations Economic Commission for Europe. Available at: http://www.unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf (accessed May 6, 2015).

UNECE and EUROSTAT. 2006. *Conference of European Statisticians Recommendations for the 2010 Censuses Of Population and Housing*. Technical Report, United Nations Economic Commission for Europe. Available at: http://www.unece.org/fileadmin/DAM/stats/publications/CES_2010_Census_Recommendations_English.pdf (accessed May 6, 2015).

Wallgren, A. and B. Wallgren. 2007. *Register-Based Statistics*. New York: Wiley.

Zhang, L.-C. 2011. "A Unit-Error Theory for Register-Based Household Statistics." *Journal of Official Statistics* 27: 415–432.

Zhang, L.-C. 2012. "Topics of Statistical Theory for Register-Based Statistics and Data Integration." *Statistica Neerlandica* 66: 41–63. Doi: http://dx.doi.org/10.1111/j.1467-9574.2011.00508.x.

Received July 2013
Revised April 2014
Accepted May 2014