ESSnet KOMUSO

Quality in Multisource Statistics

https://ec.europa.eu/eurostat/cros/content/essnet-quality-multisource-statistics-komuso_en

Specific Grant Agreement No 3 (SGA-3)

Framework Partnership Agreement Number 07112.2015.003-2015.226

Specific Grant Agreement Number 07112.2018.007-2018.0444

Work Package 3
Quality measures and indicators

Complete Overview of Quality Measures and Calculation Methods (QMCMs)

Version 2019-06-25

Prepared by Ton de Waal, Arnout van Delden and Sander Scholtus (Statistics Netherlands)

Reviewers:
Gabriele Ascari (ISTAT), Giovanna Brancato (ISTAT), Thomas Burg (Statistics Austria), Danute Krapavickaite (Lithuania), Coen Hendriks (SSB, Norway), Li-Chun Zhang (SSB, Norway), Angela McCourt (CSO, Ireland), Peter Doherty (CSO, Ireland)

ESSnet co-ordinator: Niels Ploug (DST, Denmark), email npl@dst.dk, telephone +45 2033 0875

Contents

1. Introduction

In this report, we give a complete overview of the products of Work Package 3 ("Quality measures and indicators") of Specific Grant Agreement (SGA) 3 of *Komuso* (ESSnet on quality of multisource statistics). The products are a set of output quality measures, supported by applications and computation details, that will integrate and complement the ESS Quality Guidelines for Multisource Statistics, where they are referred to and provided as an annex. Those ESS Quality Guidelines for Multisource Statistics will be produced by Work Package 1 ("Guidelines on the quality of multisource statistics") of the same ESSnet. The products may be updated later in SGA 3 in order to take into account feedback on the current versions.

In "Guidelines on the quality of multisource statistics", multi-source statistics are defined as statistics produced using several complementary data sources. The source data can range from survey data (sample or census), administrative or any other kind of data obtained from public or private data owners. In single-source statistics, one may also use additional data sources besides the main data source. A fundamental difference between multi-source statistics and single-source statistics is that, in single-source statistics, those additional data sources are used in auxiliary processes only, for instance to validate results based on the main data source. In the current document, we will use the same definition of multi-source statistics. In "Guidelines on the quality of multisource statistics" and in the current document, the combination of sources is restricted to statistical surveys and administrative datasets. Examples of so-called 'big data' are excluded.

The aim of the ESSnet is to produce usable quality guidelines for National Statistical Institutes (NSIs) that are specific enough to be used in statistical production at those NSIs. The guidelines cover the entire production chain (input, process, output). They aim to take into account the diversity of situations in which NSIs work and the restrictions on data availability. The quality of the final output will depend both on the existing data sources and on the use and processing of the data. It is, therefore, clear that general decision rules and single thresholds do not suffice. Instead, the guidelines list a variety of potential indicators/measures, indicating for each their applicability and in what situation it is preferred or not and provide an ample set of examples of specific cases and decision-making processes.

For this reason, the first SGA of the ESSnet identified several 'basic data configurations' (BDCs) for the use of administrative data sources, in combination with other sources, for which it proposed, revised and tested some measures for the accuracy of the output (see Komuso, 2017). The second SGA Work Package 3 of the ESSnet continued this work on indicators/measures, by developing further quality indicators/measures related to process and output needed for the use in practice of the guidelines. In particular, we documented the examined quality indicators/measures and accompanying calculation methods. These documents are referred to as 'quality measures and calculation methods' (QMCMs). In SGA 3, we complete the work on these QMCMs. The QMCMs, and related hands-on examples, are the products of Work Package 3 that form the Annex to the 'Quality Guidelines for Multisource Statistics'.

Until the end of SGA 3 (mid-October 2019) we will, where considered necessary or useful, update the produced QMCMs and related hands-on examples. For example, in the current deliverable we have added keywords to all QMCMs in order to ensure that relevant QMCMs can be found by (potential) users of the QMCMs. We also give a brief overview of commonly used methods for measuring output quality for multisource statistics.

The remainder of this report is organized as follows: Section 2 lists the QMCMs and related hands-on examples that have been developed in Work Package 3; Section 3 briefly describes these QMCMs and Section 4 summarizes commonly used methods for measuring output quality for multisource statistics.

2. Overview of QMCMs and examples[1]

Below we present quality indicators/measures for which we have produced QMCMs.

In order to cover different situations of different NSIs, and for ease of finding the results, we have structured the quality measures along four classifications:

- quality dimensions
- BDCs
- error types
- computational methods

The first three classifications are used in the current section. The last classification based on computational methods is discussed in Section 4. In order to improve the usefulness and ensure that the QMCMS can be easily located, we have also added keywords, and information on statistical domains, e.g. social or economic statistics, in which a QMCM is usually applied to all the QMCMs.

We focus on four quality dimensions: “Accuracy”, “Timeliness”, “Coherence” and “Relevance”. We selected quality dimensions that can be quantified. “Accuracy” is “the degree of closeness of computations or estimates to the exact or true values that the statistics were intended to measure” (Eurostat, 2014). “Timeliness” is operationalised as “the time lag between the date of the publication of the results and the last day of the reference period of the estimate of the event or phenomenon they describe” (Komuso, 2019, in line with Eurostat, 2014). “Coherence” measures the adequacy of statistics to be combined in different ways and for various uses.” (Eurostat, 2014). A more specific term than coherence is ‘numerical consistency’. Estimates, concerning different variables or different sources, are numerically consistent if they fulfil known equality and inequality constraints. “Relevance” is defined as “the degree to which statistical outputs meet current and potential user needs” (Eurostat, 2014). “It refers to whether all the statistics that are needed are produced and the extent to which concepts used (definitions, classifications etc.,) reflect user needs” (Komuso, 2019).

As already mentioned, we use a breakdown into a number of BDCs that are most commonly encountered in practice. In Komuso we have identified six BDCs. (For more information on BDCs and methods to produce multisource statistics we refer to De Waal, Van Delden and Scholtus, 2017, and Komuso, 2017):

- BDC 1: Multiple non-overlapping cross-sectional microdata sources that together provide a complete dataset without any under-coverage problems;
- BDC 2: Same as BDC 1, but with overlap between different data sources;
- BDC 3: Same as BDC 2, but now with under-coverage of the target population;
- BDC 4: Microdata and aggregated data that need to be reconciled with each other;
- BDC 5: Only aggregated data that need to be reconciled;
- BDC 6: Longitudinal data sources that need to be reconciled over time (benchmarking).

Many different schemes of error categories exist in survey and administrative sources (see, for instance, QMCM_A_2, which is discussed later). The error categories that are distinguished also depend on the level of detail that is used. Table 1 gives an overview of the error categories that are distinguished in Komuso.

[1] Part of this section is based on Van Delden, Scholtus and De Waal (2019).

Table 1. Main sources of errors in multisource statistics

Error Category	Type of error included	Survey	Admin
Validity error (VE)	Specification error (VE.SUR)	X	
	Relevance error (VE.ADM)		X
Frame and Source error (FE)	Undercoverage (FE.UE)	X	X
	Overcoverage (FE.OE)	X	X
	Duplications (FE.DU)	X	X
	Misclassification in the contact variables (FE.COE)	X	
	Misclassification in the auxiliary variables (FE.CLE)	X	X
Selection error (SE)	Error in terms of the selected sampling units (SE.SAE)	X	
	Unit non-response (SE.SUR)	X	
	Missing units in the accessed data set (SE.ADM)		X
Measurement error and Item missingness (ME)	Arising from: respondent, questionnaire, interviewer, data collection (ME.SUR)	X	
	Fallacious or missing information in admin source (ME.ADM)		X
Processing error (PE) (*)	Data entry error (PE.DE)	X	
	Coding or mapping error or misclassification (PE.CL)	X	X
	Editing and imputation error (PE.EI)	X	X
	Identification error (PE.ID)		X
	Unit error (PE.UE)		X
	Linkage errors (PE.LE)	X	X
Model error (examples, non-exhaustive) (MOE)	Editing and imputation error, record linkage error, … (MOE.PS)	X	X
	Model based estimation error (Small Area Estimation, Seasonal Adjustment, Structural Equation Modelling, Bayesian approaches, Capture-Recapture or Dual System Estimation, Statistical Matching, …..) (MOE.MB)	X	X

(*) Processing errors are errors occurring with manual activities. These include trivial errors, e.g. typographical errors in writing a procedure or errors in specifying a variable in the program (also in a model). When the processing steps mentioned are done via a model, they may result in model errors.

The QMCMs are listed in four tables: “Accuracy” (Table 2), “Coherence” (Table 3), “Relevance” (Table 4), and “Timeliness” (Table 5). In the first column of these tables, the error sources that the QMCMs aim to quantify are given. Each QMCM is given a code consisting of the letters ‘Q’,’M’,’C’ or ‘M’, a letter referring to the quality dimension (“A” stands for “Accuracy”, “C” for “Coherence”, “RV” for Relevance, and “T” for “Timeliness”), and a number. The column “Data config.” refers to the basic data configurations that have been identified in the first SGA of the ESSnet (see Komuso, 2017).

We have produced hands-on examples for all QMCMs, except for QMCM_A_2. QMCM_A_2 describes a generic framework for disentangling error types that can occur when combining survey data and administrative data. It does not provide a quality measure for combined data.

The QMCMs and related hands-on examples are made available on the CROS portal: https://ec.europa.eu/eurostat/cros/content/work-package-3-quality-measures-and-indicators-0_en.

Table 2. Measures and indicators on Accuracy

Error types	Data config.[2]	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
SE.SAE, SE.SUR, PE.EI	All	Combination of several administrative registers and/or survey datasets, at least one with missing data	Variance	The described approaches can be used when missing data isare imputed, and one aims to calculate the variance of an estimator after imputation.	QMCM_A_1 ("Sampling, non-response and imputation variance based on imputed data")	Example QMCM_A_1	Economic statistics, Social statistics	Bootstrap, imputation, multiple imputation, pseudo-population bootstrap, mass imputation, sampling error, non-response error, imputation error, categorical data, numerical data
Any	2, 3	Combination of several administrative registers and survey datasets	Bias, variance and validity	This is a generic framework for disentangling error sources. The development of specific estimation methods for different error sources remains. In principle, the approach can be used for any statistics. The approach has been used on employment status production process.	QMCM_A_2 ("Modelling of total error in multisource statistical data")	No example	Economic statistics, Social statistics	Two-phase life-cycle data model, total error framework, metadata, exert opinion, measurement dimension, representation dimension, validity error, measurement error, processing error, frame error, selection error, non-response error, relevance error, mapping error, comparability error, coverage error, identification error, unit error, categorical data, numerical data
SE.SAE, ME	5	Combination of aggregated data	Variance	The approach can be used to estimate the variance of reconciled totals and the	QMCM_A_3, QMCM_A_4 ("Variance-covariance	Example QMCM_A_3, Example QMCM_A_4	Economic statistics,	Reconciliation, macro-integration, restricted microdata, variance-covariance matrix,

[2] The data configurations are based on the basic data configurations identified in SGA 1 (see Komuso, 2017). "1" means complementary microdata sources, "2" overlapping microdata sources, "3" overlapping microdata sources with under-coverage, "4" microdata and macrodata, "5" only macrodata, and "6" longitudinal data.

Error types	Data config.[2]	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
				reconciliation has done by means of a macro-integration technique. The approach has been applied to: • Reconciled data on International Transport and Trade • Small test datasets	matrix for a reconciled vector')		Macro-economic statistics "Example QMCM_A_3" focuses on Trade and Transport Statistics	sampling error, measurement error, microdata and macrodata, macrodata, numerical data
SE.SAE, ME	5	Combination of aggregated data (administrative and/or survey data)	Mean squared error	The approach has been applied to assess the quality of estimates for municipal unemployment based the Labour Force Survey.	QMCM_A_5 ("Mean squared error of small area estimates")	Example QMCM_A_5	Economic statistics, Social statistics The example focuses on the Agricultural Census	Small area estimation, basic area-level, Fay-Herriot model, Empirical Best Linear Unbiased Predictor (EBLUP), sampling error, microdata and macrodata, numerical data
SE.SAE	4	Combination of microdata and aggregated data	Variance	The approach can be used to estimate the variance of cells in tables obtained by so-called repeated weighting. The approach has been applied to the Dutch Population and Housing Census, which is based on a mix of administrative and survey data.	QMCM_A_6 ("Variance of cell values in estimated frequency tables")	Example QMCM_A_6	Social statistics The example focuses on the Structure of Earnings Survey	Repeated weighting, numerical consistency, frequency tables, calibration, sampling error, microdata and macrodata, numerical data
FE.UE, FE.OE	2	Several administrative or survey sources with overlapping units and variables	Bias, variance	The approach has been applied to the Quarterly Survey on Earnings. The approach measures accuracy of the estimates based on the predicted values.	QMCM_A_7 ("Effect of frame under-coverage / over-coverage on the estimator of total and its accuracy measures")	Example QMCM_A_7	Economic statistics, Social statistics The example focuses on the Quarterly Survey on Earnings	Under-coverage, over-coverage, changes in the population framework, ratio estimator, sampling theory, numerical (target) data, categorical (stratification) variables

Error types	Data config.[2]	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
ME	2	Combination of several administrative registers, using micro-integration	Bias	The approach has been applied to register-based employment data and Labour Force Survey data.	QMCM_A_8 ("Quality assessment of register-based outcome variable in the presence of a sample survey for the same variable")	Example QMCMA_8	Economic statistics, Social statistics The example focuses on (register-based) Employment Statistics and the Population Census	Validity error, measurement error, sampling error, multilevel model, small are model, basic area-level model, numerical data, empirical best linear unbiased prediction
FE.UE	3	Two or more (administrative) datasets	Confidence interval	The approach estimates the confidence interval for the population size and its domain size. The approach has been applied to an automated system of decentralized population registers (with information on people that are legally allowed to reside in the Netherlands and are registered as such) and a central police recognition system where suspects of offences are registered.	QMCM_A_9 ("The confidence interval for population/domain size estimator")	Example QMCM_A_9	Social statistics	Under-coverage, population size estimation, capture-recapture method, post-enumeration survey, Petersen estimator, maximum likelihood estimation, log-linear model, populations counts
ME, PE, other errors	2	Combination of several data sources with overlapping units and variables	Qualitative indicator of quality	The approach combines quantitative information with expert knowledge to compute quality indicators for the whole data editing process. Two kinds of situations are distinguished: (1) the output value comes from a data source and there are misclassifications in all data sources, or (2) the output value was imputed.	QMCM_A_10 ("Combined quality assessment indicator")	Example QMCM_A_10	Social statistics The example focuses on Labour Market Statistics	Measurement error, processing error, classification error, quality framework, external data source, Dempster-Shafer theory, imputation, categorical data, numerical data

Error types	Data config.[2]	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
				The approach has been applied to: • a register-based census • register-based labour market statistics				
PE.LE	1, 2	Two data source (administrative data and/or survey data)	Bias, variance	The approach can be used to measure the impact of linkage errors (and methods to correct for these errors) on the quality of estimated frequency tables. The approach has been applied to Census data linked to a settlement database.	QMCM_A_11 ("Variance of a bias-corrected estimator which aims to correct for bias due to linkage errors")	Example QMCM_A_11	Social statistics The example focuses on the Population Census	Probabilistic linkage, record linkage, Fellegi and Sunter theory, bias-corrected estimator, bootstrap, frequency tables, linkage error, microdata, categorical data
ME (FE.CLE)	1	Combination of administrative data and survey data (business data)	Bias, variance	The approaches examine the effect of incorrect NACE classifications in the Business Register on the quality of the output. The approach has been applied to Quarterly VAT data and survey data.	QMCM_A_12 ("Mean squared error of level estimates affected by classification error")	Example QMCM_A_12	Economic statistics The example focuses on quarterly turnover by consumers on European web shops	Classification error, misclassification, bootstrap, variance-covariance matrix, stratum totals, audit data, split population, microdata, categorical classification variables, numerical target variables
ME	2	Combination of several data sources with overlapping units and variables (categorical data)	Bias, variance	Two kinds of approaches have been studied. In one approach it is assumed that all data sources may contain errors. In the other approach it is assumed that one data source is error free and the other data sources contain auxiliary data. The approaches have been applied to: • Employment status derived from Labour Force survey	QMCM_A_13 ("Relative bias and relative mean square error")	An example is provided in ST2_1	Economic statistics, Social statistics The example focuses on the Labour Force Survey	Measurement error, longitudinal data, Hidden Markov model, latent class model, maximum-likelihood estimation, categorical data

Error types	Data config.[2]	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
				(LFS) data and administrative data • Employment status from LFS with administrative data as auxiliary variables				
ME	2	Business register (with delayed information) and survey data	Bias, variance	The approach measures the impact of the frame errors on bias and variance of the estimator of a total in the case enterprises may join, split or change their kind of activity during the year. The approach has been applied to enterprise data on turnover.	QMCM_A_14 ("Effect of stratum changes, joining and splitting of the enterprises on the estimator of a total")	Example QMCM_A_14	Economic statistics The example focuses on the Environment Protection Survey	Joining of units, splitting of units, changing a value of a classification variable after sample selection, simple random stratified sampling, two-stage sampling, sampling theory, imputation, Horvitz-Thompson estimator, numerical (target) data, categorical (classification) variable
ME	2	Two or more datasets with overlapping units and the same target subject to measurement error	Confidence intervals	The approach can be applied to measure the quality of reconciled microdata when both data sources can contain classification errors. The approach has been applied to estimate the quality of home-ownership status observed in several datasets.	QMCM_A_15 ("Variance of estimates based on reconciled microdata")	Example QMCM_A_15	Social statistics The example focuses on data on home ownership	Reconciliation, latent class model, multiple imputation, bootstrap, measurement error, microdata, categorical data
ME	2, 6	Combination of several longitudinal data sources with overlapping units and variables (categorical data)	Misclassification rate	The approach measures the misclassification rate for observed variable with respect to target variable. The approach has been applied to: • administrative data and survey data on home-ownership • register data and survey data on jobs and benefits	QMCM_A_16 ("Misclassification rates of observed categorical variables in longitudinal data")	Example QMCM_A_16	Economic statistics, Social statistics The example focuses on the Labour Force Survey and on data on social benefits	Longitudinal data, misclassification, hidden Markov model, latent class model, measurement error, microdata, longitudinal data, categorical data

Error types	Data config.[2]	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
FE.COE, PE.LE	2	Several administrative datasets	Aggregated predicted person–place probabilities for housing units	The approach can be used to assess the effect of classification errors on the output. To this end, an assessment of the so-called ROC curve (plot of the true positive rate against the false positive rate) is used. The approach has been applied to census enumeration.	QMCM_A_17 ("Aggregate predicted person-place probabilities for housing units')	Example is provided in LR2_4	Social statistics The example focuses on Census enumeration	Contact address error, linkage error, predicted matching probabilities, logistic regression, random forests, address variable
VE, ME	2	Combination of administrative data with survey data, with overlapping units and variables (numerical data)	Validity of observed variable as indicator for target variable, Bias due to measurement error	The approach estimates the effect of measurement errors in administrative and survey variables by structural equation models. The approach has been applied to VAT data and survey data for turnover.	QMCM_A_18 ("Validity and measurement bias of observed numerical variables as indicators for a target variable")	Example QMCM_A_18	Economic statistics The example focuses on the Structural Business Statistics and VAT data	Validity, validity error, structural equation model, validity coefficient, measurement bias, audit data, measurement error, microdata, numerical data
ME (FE.CLE)	1	Combination of longitudinal administrative data and survey data	Bias, variance	The approach derives analytical expressions for the accuracy of growth rates as affected by classification errors. The approach has been applied to quarterly turnover growth rates based on business register and survey data (short term statistics).	QMCM_A_19 ("Bias and variance of growth rates affected by classification errors")	Example QMCM_A_19	Economic statistics The example focuses on turnover growth	Growth rate, misclassification, bootstrap, Markov model, stratum estimates, classification error, split population, microdata, categorical classification variables, numerical target variables, longitudinal data
FE.CLE, ME.ADM	3	Combination of two administrative datasets, or combination of survey data and	Bias, variance	The approach derives analytical expressions for bias and variance for a domain total due to under-coverage or non-response in a classification variable.	QMCM_A_20 ("Effect of frame under-coverage of a classification variable on the domain	Example QMCM_A_20	Economic statistics, Social statistics The example focuses on a survey on	Under-coverage, non-response, simple random sampling, subpopulation domain total, subpopulation domain mean, survey sampling, numerical (target) data,

Error types	Data config.[2]	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
		administrative data			estimates of the total in social statistics”)		income and living conditions	categorical (classification) variable
Any	6	High frequency times series in combination with low frequency times series	Covariance matrix of reconciled data; Quadratic function of differences between initial and reconciled data	High frequency time series are often reconciled with low frequency time series. The quality of the reconciled data can be measured by means of the covariance matrix of the reconciled time series, and by a quadratic function based on the (relative) differences between the initial time series and the reconciled time series.	QMCM_A_21 (“Macro integration: data reconciliation”)	Example QMCM_A_21	Economic statistics	Time series, benchmarking, reconciliation, quadratic optimization, linear restrictions, generalized least-squares method, Stone’s method, univariate Denton method, multivariate Denton method, Cholette adaptation, numerical data
FE.UE, ME	3	Multiple population-size statistical registers	Interval estimate	Census-like statistics can be produced by direct tabulation from a single data file, obtained from combining multiple statistical registers. To assess the accuracy of the resulting statistics in the absence of any data that are purposely designed and collected, one can model the underlying input registers and carry out estimation under the working statistical model.	QMCM_A_22 (“Accuracy of multisource census-like statistics”)	A brief example is provided in the QMCM itself	Social statistics The example focuses on Census(-like) statistics	Under-coverage, measurement error, prediction interval, interval estimate, statistical modelling and estimation, clustered capture-recapture data, dual system estimation, register-based statistical production, bootstrap, log-linear model, categorical data, population size estimation
FE.UE,FE.OE, ME	3	Household Register and a sample of households	Interval estimate	Census-like household statistics can be produced by direct tabulation from a statistical Household Register (HR). The households created in the HR have errors if people from different households are grouped into the same HR-household	QMCM_A_23 (“Household unit error in a statistical register”)	A brief example is provided in the QMCM itself	Social statistics The example focuses on the Housing Register	Under-coverage, over-coverage, measurement error, statistical household register, unit error theory, survey sampling, expert auditing, interval estimate, statistical modelling and

Error types	Data config.[2]	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
				(HRH), or if the people in the same household are divided into different HRHs. To assess the accuracy of household statistics based on the HR, one can apply the unit error theory.				estimation, prediction interval, categorical data
ME	2	Economic data, e.g. from the annual Structural Business Statistics	Posterior error probabilities; Variance inflation	A target variable is observed in several data sources and the sources cover only subsets of the target population. The "true" values of a numeric variable are predicted by means of a model. The presence of errors in a source is modelled as a Bernoulli variable; its parameter measures the error probability for the source. The observed values are characterized by measurement error, whose variance is inflated by a constant term that can be considered as an indicator of the effect of the error on the observed data.	QMCM_A_24 ("Error probabilities and variance inflation due to measurement error")	Example QMCM_A_24	Economic statistics, Social statistics	Multi-source data, data integration, contamination models, latent variables, numerical data, maximum likelihood estimation

Table 3. Measures and indicators on Coherence

Error types	Data config.	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
SE.SAE, ME	5	Combination of microdata and aggregated data	Scalar uncertainty measure	The approach measures the uncertainty in reconciled estimated accounting equations. The approach has been applied to quarterly and annual supply and use tables.	QMCM_C_1 ("Scalar measure of uncertainty in economic accounts")	Example QMCM_C_1	Economic statistics, Social statistics	Reconciliation, accounting equations, additive constraints, multiplicative constraints, Monte Carlo algorithm, sampling error, measurement error,

Error types	Data config.	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
								macrodata, numerical data
Any	All	Any type of data sources, where an external source is available	Indicators on cross-domain coherence	In this approach, released data are compared to estimates from other sources.	QMCM_C_2 ("Cross-domain and sub-annual vs annual statistics coherence")	Example is provided in ST_C_4	Economic statistics, Social statistics The example focuses on unemployment statistics	Relative difference between estimates, magnitude of discrepancy between estimates, comparison data, comparison estimate, numerical data, categorical data
Any	5	All kinds of data sources	Mean Absolute Revision (MAR); Relative Mean Absolute Revision (RMAR); Mean Revision (MR)	Application to real data in which a revision indicator, used to measure "reliability", is also used to measure the coherence among several related datasets.	QMCM_C_3 ("Cross-domain and sub-annual vs annual statistics coherence")	Example is provided in the QMCM itself	Economic statistics, Social statistics	Revision, revision indicators, seasonal adjustment, trade-off between timeliness and accuracy, categorical data, numerical data
Any	6	Longitudinal data obtained from different sources, or sets of indicator values obtained from different sources	Differences, relative differences and ratios at fixed time points; Root mean squared distance and ratio of two totals over time	Discrepancies in the values of two time series at fixed time points in their size, relative size and ratio are proposed to be measured by individual measures. Average discrepancies in two time series over time are proposed to be measured by aggregated measures: root mean squared distance and ratio of two totals.	QMCM_C_4 ("Elementary coherence measures for time series: differences, relative differences and ratios at fixed time points; root mean squared distance and ratio of two totals over time")	Example QMCM_C_4	Economic statistics	Discrepancy, time series, congeneric time series, individual coherence measure at fixed time, aggregated coherence measure over time, longitudinal data, sets of values over the same domains, numerical data
Any	6	Time series and longitudinal data from sample	Correlation coefficient and coherence	The correlation coefficient shows the degree of the linear	QMCM_C_5 ("Correlation coefficient and	Example QMCM_C_5	Economic statistics, Social statistics	Time series, longitudinal data, correlation coefficient of time series,

Error types	Data config.	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
		surveys and administrative data sources	coefficient for time series	dependency between two time series. The squared estimated correlation coefficient of two time series shows the degree of dependency among the time series. The coherence coefficient measures the degree of linear dependency of two time series indexed by the same frequency components.	coherence coefficient for time series")		The example focuses on the Labour Force Survey	coherence coefficient of time series, Fourier transformation, Fourier frequencies, periodogram, numerical data
SE.SAE, ME	5	Combination of estimates for population counts with information about demographic events	Differences between observed and reconciled data	Population count estimates (stocks) and demographic events (flows) from civil registries due to sampling and non-sampling errors might be incoherent, that is, they do not satisfy the demographic balancing equation. Usually, these stocks and flows are then reconciled. Several indicators are proposed for measuring the coherence between the observed and reconciled data.	QMCM_C_6 ("Scalar measure of coherence in reconciled balancing demographic equation")	Example QMCM_C_6	Social statistics The example focuses on demographic data	Stocks, flows, population counts, demographic events, demographic balancing equation, incoherence, balancing, Stone-Byron, accounting equation, demographic data, macro-economic data

Table 4. Measures and indicators on Relevance

Error types	Data config.	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
VE.ADM	All	Administrative data only	Expert opinion based on questionnaire	The quality assessment tool described in the paper is based on a set of questions to be answered by the statistical agency on the one hand, and by the administrative agency on the other. In particular, the statistical	QMCM_RV_1 ("Questionnaire with open questions")	Example is provided in LR0_1	Economic statistics, Social statistics	Sampling error, coverage error, measurement error, non-response error, processing error, expert opinion, questionnaire, quantitative scores,

Error types	Data config.	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
				agency has to document their quality expectations to the data source whereas the data supplier has to document the actual characteristics of their data or processes.				quality assessment tool, quality expectation, categorical data, numerical data

Table 5. Measures and indicators on Timeliness

Error types	Data config.	Sources	Quality measure / indicator	More information	QMCM	Example	Domain	Keywords
SE	2	Combination of several administrative	Bias	Effect of progressiveness (delayed input data). The approach has been applied to employment status data.	QMCM_T_1 ("Effect of delay on output estimates")	An example provided in ST2_7	Economic statistics, Social statistics The example focuses on data on employment status	Delayed inclusion into input registers, time lag, reference time point of statistics, reference time of reporting, bias modelling, progressive statistical data, t-test, confidence interval, numerical data

3. Descriptions of QMCMs

In this section, we give brief descriptions of the QMCMs mentioned in Section 2. The descriptions are largely based on the description of these QMCMs in 'Quality Guidelines for Multisource Statistics' produced by Work Package 1 of the ESSnet (see Komuso, 2019).

3.1 Accuracy

QMCM_A_1: Sampling, non-response and imputation variance based on imputed data
QMCM_A_1 describes the frequently occurring situation where missing data are imputed, and one aims to calculate the variance of an estimator after imputation. QMCM_A_1 discusses three computation methods. Depending on the specific situation, one of these three methods may be selected:

- Multiple imputation based on a prior for the model parameters;
- A direct bootstrap procedure, followed by re-imputation of the data;
- Multiple imputation based on a bootstrap procedure for the model parameters.

QMCM_A_1 pays special attention to the situation in which a sample is drawn by a complex sampling design from the population, rather than by means of simple random sampling or when the sampling fraction is not negligible. A specific example of this situation occurs when a sample survey containing the variable of interest y is linked with an administrative dataset not containing y that covers the entire population, and we impute y for all population units in which y is not observed. This is called mass imputation. Such a situation occurs in several EU member states that construct a "virtual" population census that consists of administrative datasets that cover the entire population, combined with one or more sample surveys. For instance, the variable "highest education level attained" is observed for part of the population only in most EU countries. In some EU countries, this variable is combined with administrative data that are available for all population units, and "highest educational level attained" is imputed for every population unit in which its value was not observed.

QMCM_A_2: Modelling of total error in multisource statistical data
QMCM_A_2 describes a two-phase framework to measure the total error for the situation in which data from multiple sources are integrated to create a statistical micro dataset. This total error comprises the errors occurring during the construction of each input source (first phase) and the errors occurring during the integration process (second phase). QMCM_A_2 suggests obtaining at least qualitative descriptions for the envisaged errors and their nature. Quantitative indicators depend on the nature of the errors and the availability of a target dataset and the statistical knowledge and metadata needed for the computation.

QMCM_A_3 and QMCM_A_4: Variance-covariance matrix for a reconciled vector
An issue that may arise when using multiple sources directly for producing estimates is that the estimates may differ and require reconciliation. This problem, from a macro-data point of view, has been studied in QMCM_A_3. The proposed methodology, applied by the authors to international trade and transport statistics, considers that, besides the initial set of estimates of the variables of interest, there may also be additional information on these variables derived by other sources and linear equalities between these variables that should hold by definition. The additional information

and its covariance matrix can be used to update the initial estimates and variance matrix. This leads to reconciled estimates and an updated variance matrix. Obviously, the smaller the updated variances, the more accurate the reconciled estimates can be considered, as the updated variance estimates can be considered a quality measure for the accuracy after the reconciliation process.

If the variables of interest are also subject to inequality restrictions, the methodology described should be integrated with the 'border method' or the 'approximate moments' method to obtain the vector of reconciled variables estimates. In QMCM_A_3, data are assumed to follow a normal distribution and initial estimates should be available for all variables of interest, albeit not necessarily from a single source.

Another application dealing with the same problem (QMCM_A_4), but within a quadratic programming framework, has also been developed. Here, the assumptions which inform the model are similar and the calculations to be computed depend on the type of restrictions that have to be obeyed.

QMCM_A_5: Mean squared error of small area estimates

Small area estimation is a technique aimed at improving the accuracy of estimates for samples involving small sections of the target population. This methodology is based on the use of auxiliary data and on some assumptions regarding the input estimates and the distribution of the errors. If such assumptions do not hold true, the final small area estimates may be biased. QMCM_A_5 illustrates the application of small area estimation methodology on municipalities.

QMCM_A_6: Variance of cell values in estimated frequency tables

In some cases, administrative datasets covering the entire population have to be combined with several sample surveys. When one aims to have numerical consistency between estimates of the combined dataset, one then needs to align the estimates obtained from a survey sample not only with the administrative datasets, but also with the other survey samples. This is, for instance, the case for the Dutch Population Census. To handle this complicated situation, repeated weighting (RW) has been developed as a technique for obtaining numerically consistent estimates for a mix of administrative datasets and survey samples. The RW procedure allows one to combine multiple sample surveys, instead of only one sample survey, with administrative data. To assess the quality of the estimates obtained by the RW procedure, one can compute their sampling variance. QMCM_A_6 discusses how this can be done.

The RW procedure basically consists in repeated applications of the calibration estimator, where results are calibrated on previously estimated figures. In particular, the variance-covariance matrix of the RW estimator, whose diagonals provide estimated variances for the frequency tables, is the product of 'super-residuals', linear combinations of ordinary residuals. These computations can get quite complex since the calibration is based on already estimated figures.

QMCM_A_7: Effect of frame under-coverage / over-coverage on the estimator of total and its accuracy measures

It is often the case that an administrative source is updated constantly for non-statistical purposes and thus ends up more complete than a traditional frame, such a sampling list or a fixed register. In situations like this, the frame that is used for sampling will probably be affected by under-coverage or over-coverage, for it does not take into account the units that have entered or exited the population. So, it may be a legitimate choice to adopt the complete administrative archive as an auxiliary source. QMCM_A_7 studies the effect of coverage issues on the business register frame and the relation with the social insurance inspection database, characterised by perfect coverage. The database contains an auxiliary variable, correlated with the variable of interest that can be used for the computation of the estimator when the changes in the population size are considered.

QMCM_A_8: Quality assessment of register-based outcome variable in the presence of a sample survey for the same variable

If a variable can be estimated from an administrative-based register, bias of the estimate not only gives an indication of the accuracy of the estimator, but may also indicate invalidity. In the context of QMCM_A_8, it is impossible to distinguish validity error from measurement error. In order to assess the validity error, along with the overall bias of the register-based statistic, QMCM_A_8 proposes a comparison with an estimate of the same variable from a survey, if such a source is available. The proposed quality measure is the estimated bias of the register-based subpopulation estimator. This bias is estimated by applying a small area model on the survey data in combination with the register data. The proposed quality measure is a weighted average of the directly observed difference between the register-based and survey-based subpopulation estimates, and the average bias of the register-based estimator across all subpopulations. Assumptions of the proposed quality measure are no variance for the register-based estimator and no bias for the survey-base estimator.

QMCM_A_9: The confidence interval for population/domain size estimator

In the presence of under-coverage in a frame, not only the estimates of the population variables are affected by greater inaccuracy, but the population size itself is subject to uncertainty. Therefore, it is useful to obtain a confidence interval for this latter quantity, if this quantity is estimated. The method proposed in QMCM_A_9, which is based on capture-recapture techniques, assumes the presence of two separate lists, one from the population census, the other from a post-enumeration survey. The method is based on four main assumptions: (i) independence of inclusion (the probability of inclusion in one list is independent of the probability of inclusion in the other list), (ii) inclusion probabilities are homogeneous for at least one list, (iii) the population is closed, and (iv) elements in the two lists can be perfectly linked.

QMCM_A_9 proposes the use of the confidence interval estimate for the population/domain size as a quality measure for the estimated population/domain size. The narrower the confidence interval is at a given confidence level, the more accurate the population size estimate is. The analysis can also be generalised to the case of three data sources for which the computation of parametric bootstrap confidence intervals for the population size is suggested.

QMCM_A_10: Combined quality assessment indicator

Using multiple data sources to generate statistics needs several process steps. The framework described in QMCM_A_10 introduces an indicator between 0 and 1 assessing the quality in every stage of the data processing (raw administrative data; the combined dataset, i.e. the integration of registers; and the final dataset, i.e. after imputation of missing data) for each attribute. Due to the modular design, every step of the framework could be applied individually. The approach for the assessment of administrative data relies on four quality-related hyper-dimensions (documentation, pre-processing, external sources and imputations). Documentation describes quality-related processes as well as the documentation of the data (metadata) at the administrative authorities. The degree of confidence and reliability of the data source keeper was monitored by using a questionnaire. Pre-processing refers to the proportion of data records that cannot be used. In the external source dimension the administrative data source is compared with another source, for example the labour force survey, by matching individual records and computing the share of consistent observations per variable and administrative data source. The entire information from the registers is combined with the central database which covers all attributes of interest. At this level, a quality indicator for each attribute across all data sources is computed. If a variable is only derived from one administrative data source, then the quality of this attribute on raw data level is the same as in the central database. If several administrative data sources are combined in order to derive a variable or to establish the most plausible value, then the quality indicator is calculated. This is done by using the Dempster-Shafer theory in order to combine quality indicators from different data sources. In addition, a comparison with an external source is carried out. In the last step of the data processing, missing values in the central database are imputed. For the assessment of the data quality in the final dataset, the quality indicator for imputation is computed.

QMCM_A_11: Variance of a bias-corrected estimator which aims to correct for bias due to linkage errors

When two datasets are linked through a non-unique identifier, errors may occur and the resulting estimates may be biased. QMCM_A_11 deals with this issue by adopting a probabilistic linkage procedure and computing an estimator aiming to correct for the linkage error bias. The variance of this estimator is a measure of the accuracy of the estimates. The probabilities of linkage errors are grouped in a matrix, which enters in the computation of the bias-corrected estimator and its variance. The variance itself is made up of three components, two of which can be estimated through a bootstrap procedure and the third analytically. One of the main advantages of this method is that it can be applied to more than one probabilistic linkage model. On the other hand, assuming a homogeneous distribution of the linkage errors probabilities may be violated in practice.

QMCM_A_12: Mean squared error of level estimates affected by classification error

In QMCM_A_12, stratum estimates are obtained by adding up the data within each stratum. However, the variable on which the division of the strata depends is affected by classification errors. This leads to errors in the stratum totals. Classification errors are described by a transition matrix, containing classification errors probabilities estimated through an independent and error-free collection of data. Once such probabilities have been estimated, the following step concerns the assessment of the bias and variances through a bootstrap procedure (when dealing with level estimates of stratum totals, analytic formulae can be used instead). The main obstacle in this method may be obtaining a sample of data for the estimation of the transition matrix that are clean of classification errors.

QMCM_A_13: Overlapping categorical variables without a benchmark: Integration of administrative and survey data through hidden Markov models for the production of labour statistics

QMCM_A_13 proposes, in a model-based approach, a measure of accuracy with respect to measurement error, in particular, the error in classifying individuals with respect to employment status. Contrary to classical approaches, this method entails an unsupervised approach to the use of administrative data along with a traditional survey sample. This is done by considering the target variables as latent variables, of which researchers can only obtain imperfect measures.

The application has been used on Italian market labour data, specifically from the labour force survey and related administrative data. Data is used to draw g estimates for a target variable, where g is the number of available sources. Such estimates are part of the measurement model, which may be affected by measurement errors and model misspecification.

Data is modelled following 'hidden Markov models' and estimates are obtained through likelihood methods. Simulations are also carried out by the authors to assess the robustness of the methodology with respect to departures from the model assumptions. Furthermore, distributions of the model parameters can be used to assess the quality of each source.

QMCM_A_14: Effect of stratum changes, joining and splitting of the enterprises on the estimator of a total

QMCM_A_14 concerns the changes, and the consequent measurement error, that may occur in a sample after the units have been selected. Changes may be acquired with delays in a register, resulting in temporarily incorrect information. For example, in a business population, in a stratified sample design, some businesses may be assigned to a wrong stratum due to changes in their number of employees.

Specifically, QMCM_A_14 focuses on three types of measurement errors deriving from delayed information: errors due to sampling units joining; errors due to sampling units splitting; and errors caused by changes in a classification variable. The three errors are treated and measured separately. However, they all share a distinction between the selected sample and the observed sample, the latter being the sample after the changes have occurred. The total of the variable of interest in the observed sample is the quantity to be estimated. QMCM_A_14 gives analytical formulae to quantify the effect of these changes on the estimate of a population total, and hence measure the quality of the estimated population total. In particular, these formulae estimate bias and variance of the estimated population total.

QMCM_A_15: Variance of estimates based on reconciled microdata

In QMCM_A_15, a latent class model is used to produce estimates based on a register of addresses and buildings and a survey component. Estimates are computed after the reconciliation of microdata. An observed categorical variable is considered to be an expression of the true, but hidden, target variable. This application also introduces restrictions on the latent classes in order to have results that make sense logically (for example, a rent benefit receiver cannot be a home owner). The restricted model is referred to as MILC.

While the use of latent class methods can represent a benefit for the accuracy of the estimates, the drawbacks are the complications involving the calculations and the possibility of biased estimates when the covariates contain a classification error.

QMCM_A_16: Misclassification rates of observed categorical variables in longitudinal data

In a situation where two or more linked longitudinal datasets contain the same categorical variable which may be subject to misclassification, it is plausible to represent the true values (i.e. categories) of the variable at different time points by introducing a vector of latent variables. The approach adopted in QMCM_A_16 estimates, for each unit in the sources and for each time point considered, the probability that the unit belongs to the true category. The development of the latent variable through time is described by a Markov model, under the assumption that the classification errors are independent. Since this assumption may not be true in practice, various adaptations of the model are proposed, for example introducing a dependency of a classification error on certain time points, which is reasonable for the case that the data supplier repeats the same error under the same circumstances. In any case, if the assumptions are correct, the model can be used to obtain an estimation of the misclassification rates in all the variables simultaneously.

QMCM_A_17: Aggregate predicted person-place probabilities for housing units
The indicator described QMCM_A_17 is a measure of the quality of an address variable in an administrative source containing addresses information, specifically data on housing units. Thus, the application focuses on contact address errors and linkage errors, and hence potential coverage error. This has consequences for the accuracy of the estimates.

In the method, individual probabilities for each person in the housing units are computed and then aggregated. The individual probabilities convey the likelihood for a person-place combination being correct and can be calculated through various methods, including model-based ones. Then, such probabilities are aggregated (through a minimum function or a mean function), resulting in an overall indicator that ranges from 0 to 1: the closer to 1, the higher the quality of the address variable.

QMCM_A_18: Validity and measurement bias of observed numerical variables as indicators for a target variable
Information provided by one or more observed variables can be used in the estimation of correlated target variables. The quality measure proposed in QMCM_A_18 is represented by a validity coefficient given by the absolute value of the correlation between an observed variable and the target variable. The stronger the association between the two variables, the greater such coefficients will be. A value close to 1 of this coefficient indicates an absence of measurement and validity errors. In the context of QMCM_A_18, it is impossible to distinguish validity error from measurement error. A validity coefficient of less than one can be caused by systematic errors, such as validity errors due to differences in definition, or by random measurement errors.

Since the relation between an observed variable and a target variable is assumed to be linear, the measurement bias can be decomposed in a slope component and an intercept component. Both the relations between the observed and the target variables and the relations between the target variables can be described by a structural equation model that has to be estimated. Once this step is completed, the parameters of the model can be used to assess the validity coefficient, defined above. If one also wants to estimate the measurement bias, an additional random sample of the original observations is needed which can then be used as a ‘gold standard’ for the model. Alternatively, one has to assume that the data does not contain systematic measurement errors but only random measurement errors, by putting the intercept component equal to 0 and the slope equal to 1. This assumption may be reasonable only under specific circumstances, for example for some survey data.

QMCM_A_19: Bias and variance of growth rates affected by classification errors
QMCM_A_19 represents an extension of QMCM_A_12 and is based on the same concept of a classification variable with measurement errors that affect the strata definition of the population (for example, businesses stratified by NACE code with some codes being wrong). The transition matrix used in QMCM_A_12 is still used for the estimation of the classification errors in the first time point

considered. However, for the subsequent time periods another assumption is needed. Specifically, one can assume the invariability of the true values of the classification variable over time or, alternatively, their changes over time, with or without independence of the errors across different time points. More generally, a mixture of such assumptions may hold, in the sense that different assumptions can be valid at different time points. In any case, the aim of this application is the estimation of growth rates of the target variable per stratum. Their accuracy can be assessed through a measure of their bias and their variance, which in turn can be estimated either by a bootstrap method or analytically. While the bootstrap method is a straightforward extension of the one described in QMCM_A_12, for the analytical estimation different equations have to be applied depending on which of the three assumptions described above is used.

QMCM_A_20: Effect of frame under-coverage of a classification variable on the domain estimates of a total

QMCM_A_20 proposes analytical formulae for bias and variance of the estimator for a domain total due to under-coverage and/or non-response in a classification variable. The expression for the variance of the estimator for a domain total is derived under the assumption that the estimator for the domain mean in the classified subpopulation and the estimator for the domain proportion in the unclassified subpopulation are unbiased.

QMCM_A_21: Macro integration: data reconciliation

High frequency time series are often aligned with low frequency time series. For example, the sum of the figures of four quarters must be equal to the annual figure. This procedure is called ‘benchmarking’ or ‘reconciliation’. The reconciled time series of high frequency differs from the original time series. The difference between original time series and reconciled time series of high frequency constitutes an error introduced due to the data reconciliation. In order to benchmark/reconcile the data, a quadratic optimization problem with linear restrictions is often solved. QMCM_A_21 proposes to measure the quality of the reconciled data by means of the covariance matrix of the reconciled time series, and by a quadratic function based on the (relative) differences between the initial time series and the reconciled time series.

QMCM_A_22: Accuracy of multisource census-like statistics

Census-like statistics can be produced by direct tabulation from a single data file, obtained from combining multiple statistical registers. To assess the accuracy of the resulting statistics in the absence of any data that are purposely designed and collected (such as a sample survey), QMCM_A_22 proposes to model the underlying input registers and carry out estimation under the working statistical model. Instead of using the model to produce the target estimate itself, which may be risky due to potential model misspecification, one can construct a model-based prediction/confidence interval as an uncertainty measure. In this way, one can combine a register-based statistic (by micro integration) and an interval estimate (by modelling), both of which are based on the same underlying statistical registers.

QMCM_A_23: Accuracy of multisource census-like statistics

Census-like household statistics can be produced by direct tabulation from a statistical household register (HR), obtained using the methods of micro integration combining multiple relevant statistical registers. The households created in the HR have errors if people from different households are grouped into the same HR-household (HRH), or if the people in the same household are divided into different HRHs. QMCM_A_23 proposes to apply the unit error theory to assess the accuracy of the household statistics based on the HR, as well as the statistics with household as the unit, such as household income statistics. Based on a linked sample of target households, one can estimate the conditional distribution of target household given HRHs. The sample can be based on survey sampling or expert auditing. Instead of using it to produce the target statistics, one can use it to derive model-based prediction intervals as uncertainty measures. In this way, one can combine a multisource register-based statistic (by micro integration) and an interval estimate (by modelling).

QMCM_A_24: Error probabilities and variance inflation due to measurement error
QMCM_A_24 examines the situation in which a target variable is observed in several data sources and the sources cover only subsets of the target population. QMCM_A_24 presents a model for the prediction of "true" values of a numeric variable of interest, conditional on all the available information. The true values of the target variable are viewed as realizations from a latent (unobserved) variable and the distinct (possibly coinciding) observed values from different sources are considered as imperfect measurements of this latent variable. The presence of errors in a source is modelled as a Bernoulli variable. Its parameter can be used to represent the error probability for the source. These error probabilities give an indication of the quality of the data source. The observed values are characterized by measurement error, whose variance is inflated by a constant term that can be considered as an indicator of the effect of the error on the observed data.

3.2 Coherence

QMCM_C_1: Scalar measure of uncertainty in economic accounts
For QMCM_C_1 , a number of macro variables are estimated from different sources and then reconciled to meet known accounting equations/restrictions between them. An optimization problem with restrictions is solved by the 'Lagrange multiplier method', and adjusted estimates which satisfy the accounting equations/restrictions are obtained from initial estimates. Two kinds of restrictions are studied, additive accounts and multiplicative accounts.

The proposed quality measures summarize the uncertainty of the adjusted estimates due to reconciliation. Both accuracy and coherence are estimated. Two types of measures are defined:

- The covariance approach which starts with the variance-covariance matrix as a multivariate measure of the expected deviation of the reconciled vector from its expected value;
- The deviation approach that first reduces the deviation of the reconciled vector from its expected value to a scalar summary and then defines the quality measure as the expectation of this scalar deviation measure.

The proposed measure can, for instance, be applied to a statistical system like the national accounts.

QMCM_C_2: Cross-domain and sub-annual vs annual statistics coherence
The quality indicator proposed in QMCM_C_2 is applicable when estimates of the same statistical value are available from different sources or from processes with different frequencies. The indicator is a descriptive measure computed on final estimates of the same parameter.

The relative difference between two estimates of the same parameter, computed from different sources or processes with different frequencies, can be seen as an indicator of coherence. The absolute value of the indicator provides the magnitude of the lack of coherence between the estimates, while the sign suggests its direction: a positive indicator shows an overestimation of the estimate with respect to the comparison estimate, a negative indicator shows underestimation. Reasons for incoherence have to be further explored by the researchers.

QMCM_C_3: Cross-domain and sub-annual vs annual statistics coherence

A revision is obtained as the difference between a preliminary released figure and a later calculated figure that is considered more reliable[3]. Similarly, discrepancies are obtained as the difference between an estimated figure for one domain and an estimated figure for a similar domain. Given calculated revisions (or discrepancies), QMCMC_3 proposes to compute quality indicators/measures such as the (change of) sign due to a revision, size (mean of absolute revisions, median of absolute revisions, mean of relative absolute revisions), bias (revision mean and its statistical significance, revision median) and variability (root mean square revision, range, minimum, maximum, etc.). A change of sign due to a revision indicates a potential quality problem.

QMCM_C_4: Elementary coherence measures for time series

The measures presented in QMCM_C_4 can be used to assess the coherence of two time series measuring the same indicator (for example, an initial time series with its revision) or measuring indicators under different definitions, for example the number of registered unemployed and the number of unemployed under the labour force survey definition. The measures proposed are:

- Measures at fixed time points for two time series: difference of values, relative difference of values, ratio of values for two time series show how stable coherence of these time series over time is;
- Aggregated measures over time for two time series: root mean squared distance, ratio of two totals (means) for time series over time can be used to compare the levels of coherence between several time series.

QMCM_C_5: Correlation and coherence coefficient for time series

The coherence is considered here as the extent to which the statistical outputs from different statistical processes have the potential to be reliably used in combination.

The measures can be applied for any time series that measure different indicators. They may differ because of different populations, coverage, or definitions. They may be completely different non-equivalent time series, without any possibility to decompose/aggregate them in order to obtain equivalent time series. Nevertheless, we can speak about similarity in their alternation tendencies. The tendencies of change of the time series can be compared using the following measures.

- The correlation coefficient of two time series. It shows the degree of the linear dependency between two time series. It can be positive or negative, showing positive or negative dependency between time series, correspondingly. A value of 1 means complete direct linear dependency, a value of 0 means no linear dependency at all;
- The squared estimated correlation coefficient of two time series is always positive, and shows just the degree of dependency among the time series;
- The coherence coefficient is expressed as the squared correlation coefficient of the same frequency components in the Fourier decomposition of the time series. It shows the degree of

[3] In fact, "reliability" is defined in the ESS Handbook for Quality Reports (Eurostat, 2014) as the closeness of the initial estimated value to the subsequent estimated value.

linear dependency between the components of the same frequency in the Fourier decomposition.

QMCM_C_6: Reconciling estimates of demographic stocks and flows through balancing methods
Population count estimates (stocks) and demographic events (flows) from civil registries due to sampling and non-sampling errors might be incoherent, that is, they do not satisfy the demographic balancing equation. Usually, these stocks and flows are then reconciled. An indicator measuring the degree of incoherence is the average over the regions of the differences between the direct estimate of a population count at time t and its corresponding value estimated by means of the estimates of flows and stocks of the population at time t. The sum of the differences standardised with respect to the average of the two estimates of the population count gives a global measure of data coherence. An indicator measuring the impact of the adjustments (reconciled data) on observed data is based on the relative differences between observed and reconciled data computed on each demographic figure for each region. It can be also considered as a sort of measure of incoherence, since it quantifies the number of changes of data to reach the degree of coherence. Finally, the average over the regions of the relative mean differences of the absolute values is an indicator of the impact of the procedure for data reconciliation and provides information on coherence with respect to observed data.

3.3 Relevance

QMCM_RV_1: Questionnaire with open questions
The decision to acquire an administrative source must be taken by the statistical organization on the basis of the available information regarding the source. Interaction and feedback between the organisation and the data producer should aim to reconcile the quality of the source with the expectations of the organisation . However, this process may be time-consuming and prone to errors. The tool described QMCM_RV_1 tries to solve this issue and make the communication between the two entities easier. It is a questionnaire that both the statistical administration and the data producer should answer, preferably at the beginning of the data acquisition step. The answers from both parties should clarify what is expected from the data, what uses are considered for the data, what can be done to improve future releases and so on. Besides “relevance”, other quality dimensions such as “coherence” and “reliability” are investigated. Finally, it has to be noted that no quality measures are computed in the questionnaire so it is not a quantitative tool.

For an overview of the questions in the questionnaire we refer to LR0_1 Quality Assessment Tool for Administrative Data.

3.4 Timeliness

QMCM_T_1: Effect of delay on output estimates
QMCM_T_1 illustrates that delays in data updates and transmission not only have an impact on the timeliness and punctuality of statistics, but also on their accuracy. QMCM_T_1 takes into consideration the fact that administrative registers are usually updated at different points in time, usually after an early version of the data has been collected and used in a statistical process. The more recent data, not used in the process, enjoy a better quality than the data that were used. For this reason, the resulting estimates will be affected by a bias due to the delay of the input data (delay bias).

An estimate of the average delay bias can be assessed by examining the revisions which utilise more up to date extracts. The procedure involves computing the output estimates at subsequent updates of the input data and then the differences between such estimates. At every step, the more recent estimates are assumed to be correct. The result will be a sequence of differences of which the distribution can be tested. The average value of the differences can be considered as the average bias if observations can be assumed to be independent and identically distributed, otherwise a more complex approach should be used to estimate the impact of delay bias on the quality of the statistical output.

4. Commonly used methods for measuring output quality[4]

As already mentioned in Section 2, in Work Package 3 we have made a methodological grouping of the QMCMs. In essence, one could say that there are three basic groups of computational methods: qualitative methods, numerical methods and analytical formulae. We further refined this into seven groups. The label of each group refers either directly to a methodology for estimating quality itself or to methodology for obtaining a certain kind of estimate. In the following seven subsections, we give more information and examples of such methodologies.

4.1 Using a general quality framework

The first group of QMCMs concern general quality assessment frameworks and their tooling (see Table 6). An example of such a framework is QMCM_A_10 , originally developed for the population census at Statistics Austria. Input for the population census is a large number of different administrative sources that undergo three stages of processing: the original administrative data; the combined data, i.e. the integration of the administrative data; and the final dataset, i.e. after imputation of missing data. For each of those stages, a quality indicator is computed between 0 and 1. For the quality assessment, four quality-related hyperdimensions are involved: documentation, pre-processing, external sources and imputations. The hyperdimension documentation describes quality-related processes as well as the documentation of the data (metadata) at the administrative authorities. The degree of confidence and reliability of the data source keeper was monitored using a questionnaire. Pre-processing refers to the proportion of data records that cannot be used. In the external source dimension, the administrative data source is compared with another source, for example, the labour force survey, by matching individual records and computing the share of consistent observations per variable and administrative data source. The imputation dimension computes the quality of the imputations. A total quality value per attribute and per source is computed as a weighted version of the values over the different hyperdimensions.

The datasets are combined in the central database that covers all attributes of interest. At this level, a quality indicator for each attribute across all data sources is computed. If a variable is only derived from one administrative data source, then the quality of this attribute on raw data level is the same as in the central database. If a most plausible value is derived from multiple sets, then the quality indicator is based upon the original indicator values per source, the degree to which values of the attributes in the dataset agree or not and some expert knowledge. These are combined into one value using the Dempster-Shafer theory. In addition, a comparison with an external source is carried out to capture the measurement and processing errors as well as possible. In the last step of the data processing, missing values in the central database are imputed. For the assessment of the data quality in the final dataset, the quality indicator for imputation is computed.

[4] This section is again based on Van Delden, Scholtus and De Waal (2019).

Table 6. QMCMs using a general quality framework

QMCM	BDC	Error types	Description
A_2	2,3	Any	Modelling total error in multisource data by the two-phase total error system
A_10	2	ME, PE, other errors	Accuracy-part validated with external data, combined with other error types using Dempster-Shafer-Theory
A_17	2	FE.COE, PE.LE	Assessing quality of contact addresses in administrative data by estimating probability of person-place combinations
RV_1	All	VE.ADM	A questionnaire for statistical agency and administrative agency. Part of the questions concern relevance error.

4.2 Descriptive quality measures (analytical formulae)

Next, we have a group of reasonably straightforward quality measures (see Table 7). Most of them concern the dimension "coherence" for which there are no generally agreed quality measures in official statistics (yet). One example (QMCM_C6) concerns a measure in the case of a single balancing equation. For instance, we know that the population at time t + 1 (P_i^{t+1}) of a region i should equal the population at time t (P_i^t), plus the net effect of births and deaths (N_i), plus the net population flow (immigration – emigration) M_i. Consider the situation where one has different sources for the stocks of the population and for its flows. What might happen is that the direct population estimate at time $t + 1$, $\hat{P}_i^{t+1}$, is not equal to its indirect estimate $\hat{P}_i^t + \hat{N}_i + \hat{M}_i$. Using some reconciliation procedure, reconciled estimates $(\tilde{P}_i^t, \tilde{P}_i^{t+1}, \tilde{N}_i, \tilde{M}_i)$ can be obtained. Averaged over the total number of regions (I), a measure for the effect of the reconciliation procedure, relative to the original figures is:

$$CR = \frac{1}{4I}\sum_i^I \left(\left|\frac{\tilde{P}_i^t - \hat{P}_i^t}{\hat{P}_i^t}\right| + \left|\frac{\tilde{P}_i^{t+1} - \hat{P}_i^{t+1}}{\hat{P}_i^{t+1}}\right| + \left|\frac{\tilde{N}_i - \hat{N}_i}{\hat{N}_i}\right| + \left|\frac{\tilde{M}_i - \hat{M}_i}{\hat{M}_i}\right| \right).$$

Table 7. QMCMs with descriptive quality measures

QMCM	BDC	Error types	Description
C_1	5	SE.SAE, ME	Scalar uncertainty measure derived from a matrix with (co)variances between reconciled estimates. For instance, economic accounts.
C_2	all	Any	Relative difference between the estimates of the same parameter across domains or sub-annual versus annual estimates.
C_3	5	Any	Measures for revisions or discrepancies between different statistics: change of sign, and measures of the size, the bias and variability of estimates.
C_4	6	Any	Elementary coherence measures for two time series. (Relative) differences and ratios at fixed time points and aggregations over time.
C_5	6	Any	The correlation coefficient of two times series (classical) and the (Fourier transformation) coherence coefficient of two time series.
C_6	5	SE.SAE, ME	Scalar measure for coherence with one balancing equation. The measure can be applied after balancing a demographic equation.
T_1	2	SE	Measure and t-test for the difference between estimates at two time points.

4.3 Bootstrapping

A well-known numerical method for estimating uncertainty is bootstrapping. Bootstrapping is a method of repeated sampling from either a sample (non-parametric bootstrapping) or from a distribution (parametric bootstrapping). Under certain conditions, the variance over the set of bootstrap outcomes is an approximately unbiased estimate for the variance of the original estimate. Likewise, the difference between the mean of the bootstrap estimates and the estimate derived from the original sample is an approximately unbiased estimate of the bias of the original estimate.

Examples of QMCMs that make use of this resampling technique are given in Table 8. An example at Statistics Netherlands concerns the estimation of the accuracy of highest educational level attained. This variable is derived by combining many different administrative datasets and data from the labour force survey. Unfortunately, persons born before the 1990s are underrepresented in the administrative data. One way to obtain a population estimate is to impute the highest educational level attained for the missing units in the population. QMCM_A_1 describes how the variance of an estimator as affected by imputation can be computed. In fact, three methods are explained: bootstrapping, multiple imputation and a combination of bootstrapping and multiple imputation.

Table 8. QMCMs where bootstrapping is an important component

QMCM	BDC	Error types	Description
A_1	All	SE.SAE, SE.SUR, PE.EI	Accuracy of estimates when missing data are imputed. Methods are a combination of bootstrapping and multiple imputation.
A_12	1	ME (FE.CLE)	Bias and variance of level estimates as affected by classification errors. Methods are a parametric bootstrap and analytical formulae.
A_19	1	ME (FE.CLE)	Bias and variance of growth rates affected by classification errors. Methods are a parametric bootstrap and analytical formulae.
A_23	3	FE.UE,FE.OE, ME	Bias and variance of the population distribution for number of persons in a household. The proposed method is a parametric bootstrap.

4.4 Sampling theory

In this group of QMCMs the bias and variance are computed for estimates based on random samples (see Table 9). In the case of QMCM_A_6, a large set of frequency tables is estimated (for instance for the population census) and each table is reweighted to ensure consistency with previously published tables.

Analytical variance formulae have been derived for this situation. The other three QMCMs in Table 9 refer to the impact of different kinds of frame errors in business statistics on estimators of a total. For those QMCMs multiple sources are needed to estimate the impact of those frame errors on bias and variance of the original estimate.

Table 9. QMCMs where quality measures are based on sampling theory

QMCM	BDC	Error types	Description
A_6	4	SE.SAE	Variance of cell values in estimated frequency tables that are made consistent by so-called repeated weighting.
A_7	2	FE.UE, FE.OE	Impact of coverage errors (due to use of a frozen population frame) on the variance of a separate or combined ratio estimator.
A_14	2	ME	Impact of the frame errors on bias and variance of the estimator of a total in the case enterprises merge, or change their kind of activity or size class during the year.
A_20	3	FE.CLE, ME.ADM	Bias and variance of the estimator of a total when values of a stratum classification variable are missing in one (of two) sources and their values are estimated by means of a sample.

4.5 Latent variable modelling

Consider the situation in which a variable is observed in multiple sources with some measurement error. Latent class modelling aims to model this measurement error for categorical variables and structural equation models can be used to model measurement errors in continuous variables (see Table 10 for QMCMs using latent class modelling or structural equation models). Both model types assume that there is a latent, true variable, of which the observed variables are indicators with measurement error. Additionally, true values can be fitted as a function of background variables or they can be modelled over time. In latent class modelling, one estimates the probability that a certain value is observed given the true value. In structural equation modelling, each observed value is considered to be a function of the latent true value, plus an error (see QMCM A_18). A variation is that only a certain proportion of the values per source is erroneous (see QMCM_A_24). The latter is called an 'intermittent error model' or a 'contamination model'.

Latent class modelling is, for instance, applied to the target variable, 'home ownership' (own or rent a house), measured in the Dutch LISS (Longitudinal Internet Studies for the Social Sciences) panel survey from 2013 and in a register (see QMCM_A_15). Both datasets contain the variable, 'home ownership'. Additionally, there is data on marital status and on whether someone receives rent benefit from the government. A person can only receive rent benefit if this person rents a house. In that study, the relationship between home ownership and marital status is estimated, including uncertainty estimates, while accounting for classification errors in home ownership in both sources.

Table 10. QMCMs for quality of output based on latent variable models

QMCM	BDC	Error types	Description
A_13	2	ME	With latent class modelling a 'true' value is predicted based on multiple observed categories. Resampling can be used to estimate relative bias and relative mean squared error of each source.
A_15	2	ME	Variance of output based on "true values" predicted by a latent class model is estimated by applying multiple bootstrap draws from the population, followed by one imputation from the posterior distribution.
A_16	2, 6	ME	Latent class model where the development of true value of a classification variable over time is modelled with a hidden Markov model. The misclassification rates at each time point can be computed.
A_18	2	VE, ME	Validity and measurement bias in continuous variables in different data sources measuring the same concept is estimated by a structural equation model.
A_24	2	ME	Error probabilities in continuous variables and variance inflation are estimated by an intermittent/contamination model. For simple estimators such as a total an analytical expression can be derived for the effect of ME on its variance.

4.6 Macro-integration and benchmarking

Macro-integration can be applied when one has multiple estimates at a fixed time point that have to be reconciled. QMCMs A_3 and A_4 describe how to compute the accuracy of the reconciled output, given information on the uncertainty of the original estimates. A possible application for social statistics is population census tables. Benchmarking is used to reconcile a high-frequency series to a low frequency series, which is common in economic statistics. Table 11 gives the QMCMs on macro-integration and benchmarking.

Table 11. QMCMs for quality of output based on macro-integration and benchmarking

QMCM	BDC	Error types	Description
A_3	5	SE.SAE, ME	The variance of reconciled totals estimated in a Bayesian framework where the reconciliation is done by means of macro-integration.
A_4	5	SE.SAE, ME	The variance of reconciled totals, obtained by means of macro-integration based on optimization models, estimated with sampling theory and quadratic programming theory
A_21	6	any	The variance of benchmarked data can be described by analytical formulae given the variance of the original data.

4.7 Modelling and estimation techniques

The final group of QMCMs concerns model-based estimates of which the bias or variance is estimated. Examples in the QMCMs that we describe are small area estimation methods that are used to obtain more reliable estimates for small domains (see Table 12), capture-recapture methods to estimate population sizes from multiple (incomplete) sources (see Table 13) and methods to correct for linkage errors (see Table 14).

An example of the capture-recapture approach of QMCM_A_22 is the estimation of the number of dwellings using a central population register (CPR) of persons, where the addresses suffer from measurement errors leading to underestimation of the number of resident dwellings and a building/dwelling register, which has sufficient accuracy in terms of the buildings but has under-coverage of dwellings (at the sub-building level). The approach is to link the addresses first at building level and estimate the total number (of addresses) of buildings using capture-recapture methodology. Second, one estimates the distribution of the number of dwellings per building.

Table 12. QMCMs for quality of output based on small area estimation

QMCM	BDC	Error types	Description
A_5	5	SE.SAE, ME	Mean squared error of small area estimates
A_8	2	ME	Estimated bias of a register and variance of a survey with an area level small area model

Table 13. QMCMs for quality of output based on capture-recapture methodology

QMCM	BDC	Error types	Description
A_9	3	FE.UE	Confidence interval estimation by means of the bootstrap for the population/domain size which is estimated by the capture-recapture method
A_22	3	FE.UE, ME	Confidence interval estimation by means of the bootstrap for the population/domain size in different sources, where domain sizes are estimated with a log-linear model

Table 14. QMCMs for quality of output based on probabilistic linkage

QMCM	BDC	Error types	Description
A_11	1, 2	PE.LE	Variance of a bias-corrected estimator that aims to correct for bias due to linkage errors. Combination of bootstrap with analytical formulae.

References

De Waal, T., A. van Delden and S. Scholtus (2017), Multi-source Statistics: Basic Situations and Methods. Discussion paper, Statistics Netherlands.

Eurostat (2014), ESS Handbook for Quality Reports. Available at http://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf.

Komuso (2017), *Work Package 3 Framework for the Quality Evaluation of Statistical Output Based on Multiple Sources*. Final Deliverable SGA 1 Komuso.

Komuso (2019), *Draft Quality Guidelines for Multisource Statistics – QGMSS*. Deliverable SGA 3 Komuso.

LR0_1 (2017), *Quality Assessment Tool for Administrative Data*. Deliverable SGA 1 Komuso.

LR2_4 (2017)*, A Comparison of Methodologies for Classification of Administrative Records - Quality for Census Enumeration*. Deliverable SGA 1 Komuso.

ST2_1 (2017), *Overlapping Numerical Variables without a Benchmark: Integration of Administrative Sources and Survey Data through Hidden Markov Models for the Production of Labour Statistics* (authors: D. Filipponi and U. Guarnera). Deliverable SGA 1 Komuso.

ST2_7 (2017), *Output Quality for Statistics Based on Several Administrative Sources: The Norwegian Register-Based Employment Statistics and the Effect of Delays* (author: J. Fosen). Deliverable SGA 1 Komuso.

ST_C_4 (2018), *An Applied Example of Quantifying Cross-Domain (In)Coherence on Unemployment Statistics* (author: P.J. Alexander). Deliverable SGA 2 Komuso.

Van Delden, A., S. Scholtus and T. de Waal (2019), *Methods for Measuring the Quality of Multisource Statistics*. Paper presented at the ESS Workshop on The Use of Administrative Data in Social Statistics, Valencia , 4-5 June 2019.