Ad hoc surveys: how to measure and report quality methods

Anna Di Lonardo [1], Chiara Donfrancesco [1], Laura Iannucci [2], Lidia Gargiulo [2], Luigi Palmieri [1], Flavia Carle [3], Simona Giampaoli [1]

(1) Department of cardiovascular, dysmetabolic and aging-associated diseases, Istituto Superiore di Sanità, Rome, Italy
(2) National Institute of Statistics, Rome, Italy
(3) Centre of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy

CORRESPONDING AUTHOR: *Anna Di Lonardo - Istituto Superiore di Sanità - Department of cardiovascular, dysmetabolic and aging-associated diseases Viale Regina Elena, 299 Rome, Italy - Tel 0039-(0)6-49904225, FAX 0039-(0)6-49904227 - e-mail: anna.dilonardo@iss.it*

DOI: ***10.2427/12732***

ABSTRACT

Surveys need a clear report on the quality methods adopted to support surveillance and research, and to implement evidence-based health policies. The aim of this report is to provide a simple and practical tool to the stakeholders interested in surveys, such as HIS or HES, suggesting the different aspects of quality to be controlled. This information is crucial to generate useful and reliable results that can help monitoring health policies in Europe. This report takes as its starting point the broad range of existing literature on quality methods, briefly describes the various dimensions to be documented in a quality report and how to support published results. The paper also analyses the six dimensions of quality and the different steps necessary to implement a survey, always with a special focus on quality. A survey requires that every step of its drafting be specifically planned beforehand. Moreover, training and testing of personnel involved in the different sets (organization, data collection, data processing, and statistical analysis) are fundamental elements to improve quality. The methods used and the results obtained need to be published.

Key words: health examination survey, health interview survey, quality methods

INTRODUCTION

Data quality assessment is an increasingly important element to compare health population indicators, such as frequency of diseases, high risk conditions and health performance, within and between countries. The manual of operations prepared to organize the fieldwork of ad hoc surveys should include a detailed description of the quality methods used in collecting, imputing, processing and analysing data. To understand survey results and compare them at European level, a clear report on data quality is fundamental. In preparing this report, some difficulties must be overcome, due to different and confused definitions of quality dimensions. First of all, we have to unequivocally define what is meant by quality, as we do not always refer to the same concept; different aspects of quality can be taken in consideration, depending on the context. In the case of health survey data, quality can be assessed by

measurements, missing data, data input, instruments, etc.

As it is not possible to provide a complete and exhaustive picture of all quality dimensions to be applied in an ad hoc survey, Health Examination Survey (HES) and Health Interview Survey (HIS), a review of standard quality reports was conducted to present the basic definitions of the quality components and explain how these characteristics should be reported. The aim of this paper is to provide a simple and practical tool to those interested in surveys, such as HIS or HES, and suggest the different quality aspects to be controlled. These aspects are crucial to generate useful and reliable results that can help monitoring health policies in Europe. Based on a review of international documents and our experience in the HES field, this paper aims at describing the quality dimensions of data collected through surveys and suggests how to write the quality methods chapter of the manual of operations and the data quality report that accompany the description of survey results.

METHODS

We reviewed guideline reports as well as manuals of operations adopted at national and international level for ad hoc surveys, so to have an overview of the different methodologies in use and identify common issues considered critical for quality. The official statistics quality guidelines of the Organization for Economic Co-operation and Development (OECD) [1], Canada [2], the Directorate-General of the European Commission Europe (Eurostat) [3, 4], Finland [5] and Italy (ISTAT) [6], have been selected for this revision. Moreover, we evaluated quality aspects present in the Project "Health Surveys in the EU: HIS and HIS/HES evaluations and models" [7, 8] and in HIS and HES surveys studies conducted in the field of cardiovascular diseases [9] and considered Eurostat definitions of quality in statistics [10]. A description of each dimension and of the fieldwork experience in ad hoc surveys is reported, to explain and make the different concepts of quality understandable.

RESULTS

The term "survey" defines any activity directed to collect standardized information on the general population [11, 12]. As concerns information, different surveys designs are available: a) Health Interview Surveys (HIS) [13], based on data collection through population interviews, b) Health Examination Survey (HES) based on data collection through direct examinations, standardised measures, and tests. Both HIS and HES measure biological and behavioural characteristics as well as the person's medical history, using a random sample of the general population [14, 15]. HIS surveys collect self-reported information, such as level of education, years of education, self-reported weight, height, awareness of some chronic diseases, drug treatments, smoking habits, etc.. On the contrary, HES surveys measure biological variables as blood pressure, weight, height, blood sample assay, etc.. Collected data can be analysed as crude data, using age distribution, or as age-adjusted data, using other standard age distributions. Data can be elaborated to describe variables and their associations and to generalize results by using inferential statistics.

The six dimensions of the quality

Six steps are fundamental when developing a quality report of ad hoc surveys data: 1) relevance; 2) accuracy (validity, completeness consistency); 3) timeliness and punctuality; 4) accessibility and clarity; 5) comparability; 6) coherence [16].

1. Relevance

Data analysis results should be relevant to the purposes for which they have to be used, and should meet potential users' needs. Population health and health care performance statistics are important if they significantly contribute to assess morbidity/mortality, are associated to a high utilization rate, support planning the economic resources for health systems. The importance of potentially obtainable results can change over time; therefore, requirements should be periodically reviewed, taking into account users' needs. In a quality report, relevance describes the users, the needs and how far these needs are met [16]. Questions to be answered include: 1) Who and how many are the users? 2) How important is each one of them and what are the needs that they expect to be satisfied? 3) To which extent are these needs met? Users' need may be evaluated by ad hoc satisfaction questionnaires.

2. Accuracy (validity, completeness, consistency)

Accuracy refers to the closeness of estimates to the true values. Survey results are not equal to the true values because they refer to a sample of the target population, so they are subjected to variability and bias [16]. Accuracy is a multi-faceted quality dimension because it includes different aspects that, in some cases, are interrelated with each other. Indeed, the dimension of accuracy covers validity, completeness and consistency.

Lack of accuracy can generate several kinds of errors in survey data; these errors can be distinguished in sampling and non-sampling errors. Sampling errors are due to the size of the sample population involved in

the survey; therefore results obtained by examination of a sample population are different respect to those produced by examining the whole population and errors decrease as the sample size increases.

Non-sampling errors are non-intentional errors; they are due to measurements, participation rate, coverage etc., and affect all samples. Different types of non-sampling errors can occur, and each of them can affect a different aspect of accuracy.

We describe how non-sampling errors interfere with validity, completeness and consistency.

2.1 Validity and reliability

Validity, precision, and reliability represent some aspects that contribute to accuracy. Accuracy is linked to validity, and precision is related to reliability. Accuracy, validity, precision and reliability could influence the degree to which inferences from survey results can legitimately be made.

Validity can be applied to a method or an instrument, and indicates the extent to which they measure or perform what they are designed to measure/perform. In surveys, measurement errors occur during data collection and influence validity because they generate values different from the true ones.

Measurement errors are usually generated by instruments (devices, questionnaires). For example, a laboratory method is valid when the obtained values fall within an established range. However, questionnaire errors could occur when respondents give wrong answers, consciously or unconsciously, or when interviewers influence the respondent's answer. Errors produced by instruments and interviewers can be evaluated by repeating the measurement with a different instrument, or by repeating the interview with a different interviewer. Respondent's errors are more difficult to assess, as more questions are necessary to investigate behaviours; for example, by asking two or more questions to collect a single piece of information (" Do you smoke"? and " How many cigarettes do you smoke"?). Data inconsistencies that suggest the presence of errors can be detected when editing data. The proportion of records that fail each editing is an indication of the quality of the data collection and processing.

In surveys, validity should be evaluated by content, construct and criterion validity.

Content validity refers to the extent to which a measure can evaluate the intended characteristics by comparison against standards [17, 18], to see whether measures (questions, observation logs, etc.) accurately assess what researchers want to know.

Construct validity is the process that allows to investigate the construct of a measure by means of statistical methods [19]. Constructs are abstractions that are deliberately created by researchers in order to conceptualize latent variables (not directly observed but inferred from other observed variables). Constructs are ways to indicate conditions referred to events, people, objects or things. Examples of constructs referred to people are: energy, fatigue, and disability.

Two subtypes of construct validity are known: convergent validity and discriminant validity; the former occurs when different measures are measuring the same concept, the latter occurs when a measure of one construct can be differentiated from another construct.

An example can be given by the measurement of depression (the construct) in young people. To this end, we can use two types of instruments, the survey and the observation. If the results of the two measurement methods give similar scores, we have a situation of convergent validity and, consequently, we are measuring the same construct (depression). On the contrary, if the results are different, we are measuring different constructs, (e.g., depression and anxiety) and we have a situation of discriminant validity.

Criterion validity evaluates the capability of a measure to predict an outcome. The criterion and the new measurement procedure must be related [17]. Taking again the depression example, this condition can be evaluated through the well-established 42-item questionnaire. If a shorter questionnaire has to be built, a new measurement procedure has to be created and compared against the well-established measurement method (which corresponds to the criterion validity).

Reliability is when the repetition of a method in the same conditions gives the same results. A manifestation of reliability is precision, which indicates how close the measured values are to each other. Therefore, a measure is reliable when repeatedly applied to the same population and when the same result is obtained in a high proportion of times. To assess reliability, two procedures can be applied: the test-retest procedure (the measuring procedure is performed twice on the same object, and the agreement between the results quantifies the reliability), or the inter-rate procedure (the measuring procedure is performed by two different evaluators; the agreement between the results quantifies the reliability). An example is blood pressure measurement. To obtain valid readings, it is necessary to follow standardized procedures; three consecutive blood pressure measurements are recommended, to be taken a few minutes apart, using a mercury sphygmomanometer; the measurements have to be performed on the right arm, in a quiet room with a comfortable temperature, and both measurer and patient should be in a specific sitting position. Regular calibration of the instruments increases the validity; repeated measurements could increase precision.

As mentioned, other methods to evaluate the accuracy are: the assessment of the agreement with a gold standard, as well as the agreement among data collectors and the presence of missing information. For example, the agreement of a positive questionnaire on chronic diseases assessed by medical records (gold standard) is a measure

of accuracy, while the agreement among data collectors allows to assess precision.

Missing information concerns the proportion of unknown values of variables, which indicate problems with data collection, and are due to inadequate case histories. Missing data can be categorized as: 1) Missing completely at random (MCAR), 2) Missing at random (MAR), 3) Missing not at random (MNAR) [20, 21].

Data are MCAR when the probability of missing data on a variable is unrelated to other measured variables and to the values of the variable itself. The missingness is completely unsystematic, i.e. the probability that an observation is missing is not related to any other characteristics. An example of a MCAR mechanism is a laboratory sample that can be lost, so the resulting observation (laboratory test) is missing. In this case, there is no relationship between missing observations and any other value of the examined person in the data set.

Data are MAR when the missingness is related to other measured variables, but not to the values of the variable itself. Therefore, this type of missing data is confusingly called MAR, even though there is no random missingness. For example, if men are more likely to refer their weight than women, weight is MAR. Missing values of the weight variable are not completely random, but depend on the sex variable. Another example is given by the proportion of cases with missing data, which tends to be greater amongst elderly population.

Data are MNAR when there is a relationship between the propensity of a value to be missing and its values. An example of a MNAR mechanism is the collection of information on annual income. Typically, those with higher incomes may be less willing to reveal them, so the resulting observation is missing. In this case, missing values of annual income do not depend on other observed variables, but they depend on a characteristic (higher income) of the incomplete variable itself.

2.2 Completeness

Completeness describes the degree to which values are present in the data collection. Completeness may be assessed by the number of non-available results concerning the specific issue. Completeness can be invalidated by non-sampling errors such as coverage errors and non-response errors. Before describing coverage errors, it is necessary to explain that three different levels of population exist in survey studies: the *target*, the *frame*, the *sample* populations [22, 23].

The *target population* is the population that we intend to examine; this population represents the scope of our study (e.g. the Italian adult population). The *frame population* (also called coverage) is the list of accessible statistic units (e.g. 35-79 year-old persons, resident in a selected municipality). The *sample population* corresponds to the real sample of persons extracted from the frame population (e.g. 35-79 year-old persons, resident in a selected municipality, stratified by sex and age, really participating to the screening).

The coverage errors in a survey study are caused by divergences between the target population and the frame population [16]. Ideally, the frame population coincides with the target population. Practically, this situation is difficult to obtain, because the frame population is smaller than the target population.

The following kinds of coverage errors can be distinguished:

1. Under coverage: persons not accessible by the frame; for example, residents in a given area, but temporarily out of the area;
2. Over coverage: persons accessible by the frame, but not belonging to the frame; (e.g. inclusion of dead people);
3. Multiple listings: persons present more than once in the frame (e.g. persons with two or more telephone numbers);
4. Incorrect auxiliary information: persons present with wrong information.

Coverage errors can lead to bias [16]. Over coverage, multiple listings, and incorrect auxiliary information can be avoided by checking the information about the population. Under coverage is more difficult to detect, and specialized frame quality reviews are necessary to discover this type of error.

Referring to non-response errors, these occur when persons selected in a sample are not interviewed. Lack of response can be due to different reasons: inability, unwilling, unavailability, no interest in the research, privacy problems, etc.. In the survey, non-response can involve a group of population (non-response unit) or data of some variables (non-response item) [16]. In the former, whole records are missing, in the latter, some items are missing. However, both conditions interfere with the quality of the study.

The response rate is used to quantify the extent of non-responses. It can be distinguished as unit response rate (ratio between examined and total population, with persons who answer at least one variable), and item response rate (ratio between persons that give an answer to a specific item over the total number of persons identified to respond).

The problem of non-responses is due to the fact that they introduce variability and bias in the data results. Variability is produced because the presence of non-response decreases the number of available responses.

Bias is produced because the values of the variables for non-respondents can be different from those for respondents (sailors, single persons).

Completeness of information can be evaluated by qualitative and quantitative methods. Qualitative methods estimate the degree of completeness by comparing study data to other data sources, or the same study data

over time; quantitative methods applied to HIS/HES assess the extent to which all eligible cases have been interviewed or examined (participation rate). The most known qualitative methods are historic data analyses that consist in a comparison of data with those observed in other populations that were expected to manifest similar disease rates. Differences from regional standards may reflect specific local variations in prevalence of risk factors, or the use of different methods in assessing some high risk conditions. Quantitative methods use administrative data sources to assess completeness, such as the independent case ascertainment that checks the databases of General Practitioners (GPs) to detect cases missed during HIS/HES. Examples of completeness measures of HIS/HES are the assessment of the proportion of persons examined over the eligible population (participation rate); or the incompleteness of collected information or exams (percentage of missing data for each variable) in the examined persons.

2.3 Consistency

Consistency is the plausibility degree of values within the same database or in another data set. Data consistency can be checked within a variable or between different variables (internal consistency), or at two or more points in time (historical consistency). Most quality checks performed for single variables concern format and allowed values. Taking time variables as an example, we see that if one of these variables contains information on day/month/year, it is necessary to check the range values for the day (1-31), for the month (1-12), and for the year of interest. For example, if you find 31/09/2016, there is a mistake, because this date doesn't exist. Similarly, consistency should be checked between: a) dates at birth and at diagnosis; date of birth should precede the date of diagnosis b) age, sex, and diseases/conditions: some diseases/conditions occur almost exclusively in specific age groups (children, adult, and elderly); others only in men or in women (prostate hypertrophy, menopause).

Consistency can be checked by assessing related information in cross tables (e.g., a never smoker cannot have information on current number of cigarettes/day).

Consistency can be affected by processing errors that can occur between data collection and the beginning of statistical analysis. Processing errors can be present at each single step: coding, data entry, data editing, imputation, etc.. To evaluate the impact of errors on final statistics, data should be re-coded or re-entered; disagreement between the two coding or entering phases means that some error has occurred. Error correction can be performed by checking the original questionnaire or examination, or by assigning multiple imputed values to wrong or missing data. As for measurements errors, also processing errors produce bias and variability in results of the survey.

3. Timeliness and punctuality

Timeliness refers to the length of time between the survey data collection and the availability of data analysis results. Punctuality refers to the time lag between the scheduled date, established in a calendar (reference date) and the actual delivery dates. Timeliness and punctuality refer to the frequency of released prevalence and distribution of risk factors and other results, that depend on the time needed to plan and perform the survey, and the time needed to perform quality control of data, statistical analyses, and interpretation of results [16].

Speedy access to results is a priority and a clear benefit to health providers and researchers. However, there is a trade-off between timely data and the extent of data completeness and accuracy. There is no formal definition of timeliness. It relates to the rapidity of data collection, processing and reporting of reliable and complete results. Usually, HIS data collection has a predefined time interval of 3-5 years, while HES has a 10-year time interval. A delay of some years has usually less importance in the etiological study implemented to support policy-makers to plan preventive actions than in surveys aiming at evaluating the efficacy of preventive actions or health care performance.

4. Accessibility and clarity

Accessibility is referred to the easiness of users' access to data and data analysis results , as well as the suitability of the form or media through which the information can be accessed [16]. Therefore, accessibility is related to the different aspects of dissemination, such as the distribution channel, ordering procedures, delivery, time of delivery, pricing policy, marketing conditions (copyright, etc.), availability of micro or macro data, formats (paper, files, CD-ROM, Internet, etc.).

Clarity refers to the presentation of statistics in an understandable and clear manner. Clarity presupposes that all results are accompanied by textual information, explanations, graphs, figures and other illustrations, and assistance is offered to the users by the data provider. Documents usually tend to be understandable to experts only. Efforts should be done to make metadata user-friendly also to infrequent users.

Accessibility and clarity usually are the less considered quality dimensions. If data cannot be accessible or statistics are not understandable, also the most accurate and coherent data have little value.

5. Comparability

Comparability is the extent to which differences between statistics from several geographical areas, or

over time, can be attributed to differences between the true values of the statistics [16].

In surveys, the factors responsible for loss of comparability are related to: 1) use of different definitions; 2) use of different procedures or measuring tools. Before starting a survey, it is fundamental to plan all aspects of the study related to concepts and measurements. For example, aspects to be established for the concept are: statistical unit, target and frame population, reference period and frequency, study domains, standard measurements methods, etc.. A same definition must be necessarily used when two surveys are compared. Measurement aspects include the method chosen for the measure, data collection, processing and statistical analysis. Standardised methods should be used for measurement and analysis.

Comparability of data collected in the survey can be distinguished in: 1) geographical, 2) temporal, 3) between domains, 4) combination of the above situations.

The geographical comparability refers to the comparison of similar surveys that analyse the same phenomenon, but involve the population of different geographical areas or are conducted by different organizations.

Comparability should also exist over time, but, if changes occurred, data collected in a specific reference period cannot be fully comparable with data of following periods, and consequently a break in the time series is introduced. Changes due to modification of references, concepts or measurement process should be documented, and their impact assessed.

Results from different domains can be compared, but in this case different concepts could be used (definition of characteristics, reference period, etc.). All the differences should be reported and their effects evaluated. Comparability of data is a crucial aspect to allow reliable conclusions and to perform benchmarking between countries/regions and periods.

Comparability can be ensured with proper standardization and harmonization approaches. For diseases, the basic requirement is the standardization of case definition and rules for coding and reporting multiple events when they occur in the same individual.

Precise knowledge of current and historical registration procedures, methods and definitions are of great importance in the analysis of the geographical and temporal variation. An example is reported for the history of acute myocardial infarction: in HES, this is assessed by a combination of the "London School of Hygiene and Tropical Medicine" questionnaire and some items of electrocardiogram (ECG) read by means of the Minnesota code (item 1, item 4, item 5) [24].

To assure comparability, particular attention should be given to:

a) systems used for coding: international classification of diseases version (ICD 8, ICD 9, ICD 10);

b) algorithms used for case definition (ECG and history of myocardial infarction) and the date when the disease occurred (first date of hospital admission, last date of hospital admission);

c) asymptomatic event detection: incidental detection of event (e.g. asymptomatic myocardial infarction or cancer can be detected during a HES)

In order to improve data comparability, an extensive description of methods used to collect data, process them, and produce statistics should be published and periodically updated; the description of methods should be published in journals, reports, and web-sites in order to avoid loss of adopted information on methodologies over time (web sites may change over time).

6. Coherence

The coherence of two or more data results refers to the degree to which the same concepts - classifications, definitions, and target populations – and harmonised methods were used in the statistical processes that generated these results. Incoherence is more frequent than coherence, because statistics can be generated from different sources (surveys of different fields or topics) and by using different approaches and methods.

Coherence can be easily confused with comparability, therefore it is important to underline the difference.

Both coherence and comparability refer to a dataset (HIS) with respect to another (HES). Coherence evaluates inconsistencies among actual data, while comparability is based on the use of aggregated data. In other words, coherence refers to the comparison of statistics between same or similar data collection ; whereas comparability is based on statistics from unrelated populations.

Coherence can be assessed for different areas [16].

a) coherence between interim and final statistics: a measure should be used to establish if the difference between interim and final statistics is really meaningful; a lack of coherence influences reliability;
b) coherence of annual and short term-statistics;
c) coherence of statistics in the same domain: this is when some statistics, possibly of a different type, measure the same phenomenon with different approaches.

Coherent statistics validate data each other and can be validly combined and used jointly.

DISCUSSION

Before starting a survey, it is fundamental to write and report a protocol that includes the survey's hypotheses and detailed objectives. These must be accompanied by a manual of operations describing the procedures in detail.

During the entire survey process, operations such as

planning, organization, collection, quality control analyses, interpretation and dissemination of results can help to make sure that the right questions are asked to the right target audience, so that the claimer can obtain the expected results and, more importantly, good quality data can be provided.

Survey objectives should include a general description of the information that researchers want to obtain from the survey, the target population, and an explanation of possible comparisons. The survey objectives should contain the measurement objectives, the categories of data that must be acquired, the target population that will provide these data, the possibility to compare data from previous years or with a subgroup of population, etc.. For example, the objectives of the Cardiovascular Epidemiology Observatory/Health Examination Survey (OEC/HES) were to assess the prevalence of more frequent chronic diseases and risk conditions, and assess the mean levels of risk factors and treatment indicators in a representative sample of the Italian adult population [25].

In order to practically describe how to implement the survey according to a quality approach, the planning of the following specific steps is recommended at the beginning of the work: 1) define what should be evaluated or measured; in other words, identify the measures to be used to assess the conditions in the general population (high risk groups such as hypertensive people, or dyslipidaemic people, or diabetics) - often, the approach used leads to represent specific conditions by using proportions, where the examined population represents the denominator; 2) analyse the condition/s chosen as the target of the measurement; that means that the target population shall be defined according to some selection criteria, such as age range, or prevalent cases shall be defined (i.e. only those including the confirmed diagnosis), etc.; 3) give a specific definition of the condition; 4) identify possible measurements; 5) evaluate the validity of the measures.

The manual of operations is fundamental to organize a survey [26]. It includes all specific information on methods and procedures to be adopted, such as definition of target population, sampling, measurements, questionnaires, communication to participants, data analysis and data storage. All steps included in the manual of operations should be checked during and after the survey, and quality of performance and of collected data should be tested. Country-specific conditions and health monitoring needs should be evaluated in this context. Crucial steps to improve data quality are represented by the training and testing of personnel involved in data collection procedures.

The staff performing measurements or collecting data should be qualified through training and testing. After a complete information on the study (its objectives, procedures and methods for measurements), usually given by an expert, the staff should attend practical sessions where an expert performs standardized procedures and methods for measurement and/or questionnaire administration, for results codification, and computer input; subsequently, the staff shall replicate the same procedures under the supervision of the expert. The staff is trained until it is able to perform the procedures as indicated in the manual of operations. The testing process checks the agreement of staff performance results with expected values. Training and testing can improve data reliability.

Procedures, methods and tools used during the study have an impact on the final results of the study. They should be standardized, comply with the purposes of the study, match the population under evaluation, and be completed. They should provide good quality data, do not overload the participants (in the case of ad hoc surveys), comply with ethical and data–protection requirements, and have limited costs. Rules and comments on data collection implementation should be fixed in a written form and made available to the data collection personnel [27].

Different types of errors can occur during the data collection, and can be grouped in systematic and random (casual) errors. Systematic errors occur when measurements deviate from the true value in one direction only (i.e. different laboratory methods to measure serum cholesterol), or when the completeness of data is not given for a subgroup (i.e. elderly people in a survey). Systematic errors are difficult to identify and quantify, but approaches such as sensitivity analyses could be taken into account – for instance, comparing results among subgroups or with independently assessed information.

If the systematic error is identified and the systematic deviation is assessed, the solution could be a correctional adaptation of the methods used in the study, e.g. application of the systematic deviation during the data collection or insertion of the correct values. When such solution is not applicable, the exclusion of the involved group from the analysis is recommended. Random errors can be defined as the portion of variation in a measurement or in a study that has no apparent connection to any other measurement or variable, and it is generally due to chance. This component of error is assessed through statistical methods (for examples testing standard errors of values) and could be minimized by using a good sampling method, having a good participation rate, and increasing the sample size of the study.

Quality reports represent a fundamental tool in surveys (HIS and HES), as they allow a good use and comparison of data. Quality reporting is the preparation and dissemination, on a regular or irregular basis, of reports conveying information about the quality of a statistical product or survey [16]. A quality report provides information on the main quality characteristics of a product, so that the user can assess product quality. Ideally, quality reports are based on quality indicators [4]. The "ESS Quality and Performance Indicators" represent a set of standard indicators shared at European level which should be reported in a standard quality report [28]

For this purpose, the quality report has to respect the six dimensions described above and, last, but not the least,

has to be available to all users. In conclusion, to facilitate the elaboration and writing of a quality report, a check-list with the items to be described is provided below:

1. Relevance: description of
   - □ users
   - □ needs
   - □ way
2. Accuracy (validity, completeness, consistency):
   - description of sampling errors (linked to the population under examination)
   - description of non-sampling errors:
   - □ measurements errors (validity, reliability)
   - □ coverage errors (completeness)
   - □ processing errors (consistency)
3. Timeliness and punctuality: description of
   - □ the average timeliness of data
   - □ the time lag between the scheduled date and the delivery date (punctuality)
4. Accessibility and clarity: description of conditions for access to data
   - □ Distribution network
   - □ Delivery
   - □ Marketing conditions
   - □ Form of presentation of statistics (text, graph, explanation)
5. Comparability: description of differences in statistics between
   - □ geographic area
   - □ temporal
   - □ domains
   - □ combination of the above situations
6. Coherence: description of comparison among
   - □ temporary and final statistics
   - □ annual and short term statistics;
   - □ the statistics in the same domain

Funding

The activities described in this paper were conducted in the framework of the Project "Creazione e sviluppo del Network Italiano a supporto del progetto europeo BRIDGE-Health finalizzato a dare strutturazione e sostenibilità alle attività europee nel campo della Health Information (HI) (Creation and development of the Italian Network to support the European project BRIDGE-Health aimed at structuring and providing sustainability to European activities in the field of Health Information)", funded by the Italian Ministry of Health, Centre for Disease Control (CCM), and in the framework of the European project 'BRIDGE- BRidging Information and Data Generation for Evidence-based Health Policy and Research' funded by the European Commission /DG Santè (Agreement no-664691 – BRIDGE HEALTH)

Statement

The findings and conclusions in this paper are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of European Commission/DG Santè or the Italian Ministry of Health or the Istituto Superiore di Sanità. Therefore, no statement in this report should be construed as an official position of European Commission/DG Santè or the Italian Ministry of Health or the Istituto Superiore di Sanità.

References

1. Quality guidelines for OECD statistics in Quality Framework for OECD Statistical Activities Version 2011/1.
2. Statistics Canada Quality Guidelines Fifth Edition – October 2009.
3. ESS handbook for quality reports. Eurostat Manuals and guidelines, 2014 edition.
4. Mats Bergdahl, Manfred Ehling, Eva Elvers, et al. Handbook on Data Quality Assessment Methods and Tools. European Commission, Eurostat, Wiesbaden, 2007.
5. Statistics Finland Quality Guidelines for Official Statistics; 2nd Revised Edition 2007.
6. Signore M, Brancato G, Carbini R, D'Orazio M, Simeoni G. Linee guida per la qualità dei processi statistici Versione 1.1 Dicembre 2012 Istituto nazionale di statistica- Istat.
7. Koponen P, Aromaa A. "Survey design and methodology in national health interview and health examination surveys review of literature, European survey experiences and recommendations" in Health surveys: evaluation and recommendations. Aromaa A, Koponen P, Tafforeau J et al. Publications of the National Public Health Institute, Helsinki 2003.
8. Aromaa A, Koponen P, Tafforeau J et al. Health surveys: evaluation and recommendations Publications of the National Public Health Institute, Helsinki 2003.
9. Primatesta P, Allender S, Ciccarelli P et al. Cardiovascular surveys: manual of operations. Eur J Cardiovasc Prev Rehabil 2007; Dec; 14 Suppl 3:S43-61. doi: 10.1097/01.hjr.0000277988.18096.3b.
10. Doc. Eurostat/A4/Quality/03/General/Definition Working Group "Assessment of quality in statistics" item 4.2: Methodological documents - definition of quality in statistics;
11. Czaja R, Blair J. Designing surveys. A Guide to Decisions and Procedures. 2nd edition, Sage Publications, Inc., California, USA, 2005.
12. Biemer PP, Lyberg LE. Introduction to Survey Quality. Wiley Series in Survey Methodology, John Wiley & Sons, New Jersey, USA, 2003.
13. European Health Interview Survey (EHIS wave 2) - Methodological manual 2013 edition. http://ec.europa
14. EHES Manual Part A. Planning and preparation of the survey 2nd edition Edited by Hanna Tolonen Helsinki 2016.
15. Tolonen H, Koponen P, Aromaa A, et al. (Eds.) Recommendations for the Health Examination Surveys in Europe. B21/2008, Publications of the National Public Health Institute, Helsinki 2008. Also available from http://urn.fi/URN:ISBN:978-951-740-838-7
16. Doc. Eurostat/A4/Quality/03/General/ Standard Report Working Group "Assessment of quality in statistics" item 4.2 B: Methodological documents - Standard Report.

17. Gandek B, Ware JE, Aaronson NK et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol. 1998; 51(11):1171-8.
18. Haynes SN, Richard DC, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods Psychological Assessment 1995; 7(3): 238-247.
19. Cronbach L.J. Essentials of psychological testing (5 ed) 1990 New York Harper Collins Publisher Inc.
20. Baraldi AN, Enders CK. An introduction to modern missing data analyses Journal of School Psychology 2010; 48: 5-37.
21. Donders ART, van der Heijden GJM, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values J. Clinical Epidemiology 2006; 59 (10) :1087-1091.
22. Target Population and Sampling Frame in Survey Sampling: www.theanalysisfactor.com/target-population-sampling-frame
23. International Handbook of Survey Methodology Editors E. D. de Leeuw, J. Hox, D. Dillman 2008.
24. Rose GA, Blackburn H, Gillum RF, Prineas RJ. Cardiovascular survey methods. Monograph Series. World Health Organisation 1982 Vol.56. 2nd edition pp.178.
25. Giampaoli S, Palmieri L, Donfrancesco C, Lo Noce C, Pilotto L ,Vanuzzo D on behalf of The Osservatorio Epidemiologico Cardiovascolare/Health Examination Survey Research Group. Cardiovascular health in Italy. Ten-year surveillance of cardiovascular diseases and risk factors: Osservatorio Epidemiologico Cardiovascolare/Health Examination Survey 1998–2012. Eur J Prev Cardiol 2015; 22(2S): 9–37.
26. The EHES Manual, 2nd edition (2016) : www.ehes.info/manuals.htm
27. http://dgepi.de/fileadmin/pdf/GEP_LL_english_f.pdf.
28. European Commission Eurostat, Luxembourg ESTAT/d4/la d(2014) 02-ESS Guidelines for the implementation of the ESS quality and performance Indicators (QPI).

*