BRIDGEHEALTH
BRidging Information and Data Generation for Evidence-based Health Policy and Research

HA5 - DATA QUALITY METHODS

30 October 2017

Leader: Simona Giampaoli
Istituto Superiore di Sanità
Viale Regina Elena 299, 00161 ROME, Italy
e-mail: simona.giampaoli@iss.it
tel +39 0649904231

This project is funded by the Health Programme of the European Union

CONTENTS

EXECUTIVE SUMMARY

The pursuit and maintenance of data quality in a health information system is crucial to assess population health and healthcare services, monitor time trends of diseases and geographical gradients, identify gaps and reduce inequalities.

Besides regular statistics on demography (births and deaths) and economy of the health system, the main data sources contributing to health information are: 1) administrative databases (hospital diagnoses, drug prescriptions, outpatient visits, exemptions), systematically collected at national level for the management of resources and the purposes of healthcare services; 2) health examination surveys/health interview surveys, which provide standardised data on representative samples of the general population; 3) population-based registries, which provide standardised data in definite areas under surveillance. Clear definitions, data harmonization and data processing procedures in computing indicators are the key issues to ensure reliability and comparability.

This horizontal activity aimed at: identifying quality assessment methods in data collection/data sources among previous and running European Projects, particularly in those participating to BRIDGE Health; identifying quality assessment methods in data processing from different sources to assess indicators; creating an overview of health information areas where quality issues are faced.
A core result of the Horizontal Activity is the technical report that can contribute to build an integrate, sustainable and comprehensive European (EU) health information system.

This work is based on experiences and good practices developed by experts in different European Projects; a questionnaire was sent to the Working package leaders; the literature review of the quality methods used in health data, data sources and health indicators was updated. The report includes a detailed description of quality dimensions of data and data sources (relevance, accuracy, timeliness, accessibility, comparability, coherence), a description of systematic and random errors, methods to assess quality and validity of indicators, implications and limitations, including the description of major difficulties encountered to ensure data quality in different EU projects. Examples of quality checks for data provided by *ad hoc* surveys, population-based registries and administrative databases are described, together with the main steps to improve quality methods.

The first step to plan and organize a quality data collection is to prepare a manual of operations; this manual must include a detailed description of exams/questions/data, should follow international standardised procedures and methods in definitions of the diseases under surveillance, in data collection, and in data processing. The training and testing of the personnel involved in data collection and data management ensure good quality data and reduce systematic

errors. A report with a detailed description of quality checks may help harmonise the different databases to be included in an health information system. A prompt feedback to the personnel involved in data collecting, harmonization and processing may improve data quality.

Key points

Health care delivery, resource allocation and research are severely impaired without good data, good quality of indicators and studies, and therefore decisions on planning and evaluating preventive programmes.

Quality involves different dimensions: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability and coherence. These dimensions are relevant for the use of primary data (HIS, HES, population based registries) and secondary data (administrative data)

A report with a detailed description of quality methods and checks used to assess quality data may help harmonize databases from different data sources to be included in a health information system.

One of the aims of this work was to stimulate the creation of an expert network on quality methods, so that this network could support the NCD Health Information System across Europe and create the basis for future public health policies on surveillance in all European countries.

ABBREVIATIONS AND ACRONYMS

AMI	Acute myocardial infarction
BMI	Body mass index
BRIDGE-HEALTH	Bridging Information and Data Generation for Evidence-based Health Policy and Research
CVD	Cardiovascular disease
DRG	Diagnosis related group
ECHI	European Core Health Indicators
ECHIM	European Core Health Indicators Monitoring
ECG	Electrocardiogram
EHES	European Health examination survey
EHIS	European Health interview survey
ESC	European Society of Cardiology
EU	European Union
EUROCISS	European Cardiovascular Indicators Surveillance Set
Eurostat	Statistical Office of the European Community
GP	General Practitioner
HDR	Hospital discharge record
HES	Health examination survey
HI	Health information
HIS	Health interview survey
ICD	International Classification of Diseases
IDB	European Injury Database
MDS	Minimum Data Set
MONICA	MONItoring Trends and Determinants in CArdiovascular Disease
MS	Member States
PIN	Personal identification number
PPV	Positive predictive value
WHO	World Health Organization
WHO-EUR	World Health Organization-Region of Europe

I. INTRODUCTION

The pursuit and maintenance of data quality in a health information system is crucial to assess population health and healthcare services, monitor time trends of diseases and geographical gradients, identify gaps and reduce inequalities.

Provision of healthcare services, resource allocation and research are severely impaired without good data, good quality of indicators and studies, and therefore decisions on planning and evaluating preventive programmes.

The assessment of data quality is crucial to produce reliable and comparable indicators, and therefore to build health information systems at national and European level.
Surveys, registries, and administrative databases are the main data sources that contribute to health information.

A *survey* is an investigation about the characteristics of a given population; it is used to collect standardised data from a sample of that population and estimate its characteristics through the systematic use of statistical methodology [1]. Surveys can be distinguished into Heath Examination Surveys and Health Interview Surveys.

A *Health Examination Survey (HES)* is a population-based survey conducted in a random sample of the general population of the country; data collection is based on measures and examinations that follow standardised methods and procedures (e.g. systolic and diastolic blood pressure, anthropometric measures, functional activities, electrocardiograms, spirometry), biological tests based on centralised laboratory assays (e.g. lipids, glycaemia, haemachrome), standardised questionnaire(s) (e.g. chronic diseases, life styles, pharmacological treatments, family history of diseases, diet, physical performance, cognitive function, etc.). Standardisation of measurements, training of personnel and quality control are essential to assure reliable and comparable data. It is recommended to use survey instruments whose sensitivity and specificity have already been assessed. It is also important to periodically check that instruments are perfectly functioning and that the same instruments are used for the whole duration of the survey. Personnel assigned to screening should be properly trained and quality control should be assured for the whole collection period. Several health operators may be needed to carry out a survey, and their activities should be regularly screened for quality control over the whole collection period in order to ensure validity and comparability of data. A pilot study is needed to test the entire set of procedures and methods before starting screening procedures.

A *Health Interview Survey (HIS)* is a population based survey that includes interviews on health characteristics (perceived health, diseases, disability), health

related behaviour (e.g. smoking habit, physical inactivity) and use of health services. It is based on face-to-face interviews and self-administered questionnaires, telephone interviews and postal surveys.
HIS are used to collect information on self-reported and perceived health status, health determinants and health care in samples of the general population; HES provide objective measurements of health-related outcomes, but they are expensive and time consuming. If conducted adopting proper standardised and harmonized approaches, both HIS and HES may produce comparable and reliable statistics and health indicators (e.g. prevalence).

A *Population-based registry* performs a continuous (or periodical) and systematic collection, analysis, interpretation and dissemination of information about the occurrence of a disease (e.g. cancer, cardiovascular disease), the use and monitoring of medical devices (e.g. joint replacement devices, pace-makers) or conditions (e.g. injuries) in a defined population resident in a specific geographic area (state or region). Population-based registries provide data concerning all cases of a specific disease in a defined population (residents in the area under surveillance), whether treated at home or in hospital, in whichever season of the year or time of the day they may occur, and also include fatal cases whose sudden occurrence prevents them to reach the medical services.
The strength of a chronic disease registry lies in the possibility of validating each single event (or a random sample of suspected events) according to standardised diagnostic criteria, and collecting disease-specific clinical and para-clinical data. The event identification can be obtained through a verification of the events in hospital carried out by means of *hot pursuit* or *cold pursuit* procedures. In *Hot pursuit,* cases are identified according to their admissions to hospital, usually within one or two days from event onset; relevant information is acquired by visiting the ward or interviewing the patient. In a *cold pursuit,* routine and delayed procedures are used, and cases are identified by means of hospital discharge records, review of medical records and, in case of a fatal event, death certificates. These procedures provide very high quality data. The weakness of a chronic disease registry lies in the fact that its data collection is very expensive and time consuming and these registries can usually be maintained only for a limited period of time in a defined population of a reasonable size. Registry data are usually available with a delay of 3-5 years. A pragmatic solution can be the integration of administrative databases through record linkage, in particular those mortality and hospital discharge records; this could help the identification of potential events and validation of a sub sample of events to estimate positive predictive values (PPV). PPV, applied to overall potential events, will help estimate the occurred events and, consequently, their main indicators (incidence rate, and case fatality).

Administrative data are systematically collected in local, regional, and national

databases for purposes related to management of resources, costs, and services (e.g. mortality, hospital discharge records, drug prescriptions, exemptions, etc.). They do not have a research purpose, but due to their width and richness of information they will be more and more used and interconnected with data from research studies (longitudinal studies, HES, HIS, registries), to serve research purposes.
Administrative data are attractive as they offer more advantages than population studies; some of these advantages are the low cost necessary for their collection, the large number of individuals included in these data (generally all the population), their timeliness, the long periods covered, which usually cannot be achieved through surveys. Moreover, the use of administrative data can reduce the number of questions or the burden of information to be collected by an ad hoc survey, or the need to collect data on critical information that usually individuals prefer not to provide (e.g. estimating the individual mean income without asking individuals to answer a direct question on personal income in a questionnaire). Finally, administrative data can also be used for secondary purposes as sources of information for registries, surveys or follow up of longitudinal studies by integrating data through record linkage procedures. The disadvantages of administrative data are their lack of statistical quality of since such databases are collected for administrative purposes and not for research purposes; their lack of quality control during data collection; their availability; their accessibility by researchers, and their highly limited use due to privacy and ethical issues. For all these reasons, before planning the use of administrative data (e.g. for record linkage in a population-based registry), it is necessary to weigh up the advantages and disadvantages in relation to the specific research hypothesis, and it is necessary to take into account that administrative databases must require a huge and capillary checking and cleaning.

The primary objective of population-based registries and HIS/HES is to provide information about the descriptive epidemiology of a specific health problem, such as incidence survival, prevalence of the disease in the population and its change over time. This information can be useful to evaluate prevention programmes and address health policy decisions.

A goal common to both population based registries and administrative databases is the production of relevant statistics for health care management, the planning of health services and healthcare expenditure, and the provision of data on mortality, causes of death and hospital admissions for international statistics.

Data processing harmonisation is of paramount importance to ensure comparable definitions when administrative databases are used, and when ad hoc surveys and registries are pooled.

High quality data are essential for research, to assess the health indicators used for surveillance, prevention and health care, and to support health policy makers in their activities and decisions.

II. AIMS

The Horizontal Activity on "Data quality methods including internal and external validation of indicators" (HA5) is aimed at:

- identifying quality assessment methods in data sources/data collection among previous and running European Projects, particularly in those participating in the BRIDGE Health;
- identifying quality assessment methods in the processing of data and data sources used to assess indicators;
- Creating an overview of health information areas where quality issues are faced.
-

III. APPROACH

This report is based on experiences and good practices developed in previous and current European projects and in individual Member States (MS).

A questionnaire was filled in by BRIDGE Health work package leaders, who inserted the following questions: 1) what kind of health information did/do you collect by using standardised procedures or methods? 2) Did/do you build ECHIM indicators in compliance with the recommended procedures/methods? In case you did/do not, please specify why. 3) How did/do you assess the completeness of events/information? 4) How did/do you perform internal validity? 5) How did/do you assess external validity? 6) Have you received training on standardisation procedures and methods to assess data quality? How many training sections and what type of training have you received? 7) What are the major difficulties that you have encountered to assure data quality?

A literature review of the quality methods applied in health data, data sources, and health indicators was implemented; Manuals of Operations and reports produced by EUROSTAT and the National Institute of Statistics were reviewed [2-5].

Lessons on global burden of disease learnt from EUROCISS [6] and EHES Projects [7], Non Communicable Diseases Risk Factor Collaboration and World Health Organization (WHO) [8,9], as well as from the joint experiences of fieldwork researchers involved at national level to collect data, use different data sources, process data, and assess indicators are reported.

The report includes a detailed description of data quality dimensions, data sources, and quality methods for indicators. The quality methods developed in EU Projects are also reported, with their implications and limits, including the description of the major difficulties encountered to assure data quality and the main steps to improve quality methods. Final recommendations are also suggested for quality implementation in data, data sources and indicators, and training and testing are suggested as key issues to guarantee data quality and standardisation. Examples of the quality methods applied in *ad hoc* surveys, registries, administrative databases, are also described.

This report represents a scientific support for investigators, health professionals and staff working at National Public Health Institutes, National Institutes of Statistics, Local Health Units, and other academic and public health institutions operating at both regional and national levels.

IV. DEFINITION OF QUALITY

The starting point to define quality is the definition of quality given by EUROSTAT Working Group Assessment of Quality in Statistics in ISO 8402-1986: "*the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs*" [10]. This definition was subsequently updated and improved in ISO 9000, 2005 (3.1.1 **quality:** *degree to which a set of inherent characteristics fulfils requirements*; 3.1.2 **requirement**: *need or expectation that is stated, generally implied or obligatory*) and in ISO 8000-8, 2015 *(Quality Data are data that meet* ***requirements*** *and are* ***portable*** *-defined syntax and defined semantic encoding-*)

It is difficult to provide the exact meaning of quality; when we speak of quality, we do not always refer to the same concept. Different aspects of quality (dimensions) can be taken into consideration according to the various contexts, disease/characteristics, type of data, data sources and indicators.

A **case/event definition** is a set of standard criteria used to the presence of a particular disease, syndrome, or other health condition in a patient. A standardised epidemiological definition of event should be adopted and should take into consideration, as far as possible, specific operative conventional agreements. A case definition must be clear, simple, and concise, so to be easily applied to all individuals in the population of interest. It can be defined by clinical and laboratory characteristics obtained through various methods - such as for instance a diagnosis by a physician, the completion of a survey, or routine population screening methods. To assure data comparability, it is crucial to use always the same case definition.

Data means those characteristics or information, usually numerical, collected through observations [11].

Data sources are specific datasets, metadata sets, databases or metadata repositories where data or metadata are available [12]. According to the various ways in which data are collected, data sources can be distinguished in administrative, survey (HIS, HES) and registry sources[12]

Indicators are summary measures related to a key issue or phenomenon and are deduced from a series of observed facts (data). Indicators are useful to identify trends and draw attention to particular issues. They are necessary to monitor disease and health, set policy priorities and benchmarking [11].

Methods to assess quality concern data, data sources and indicators. Quality is a multidimensional concept and cannot be duly assessed by using a single measure only. The quality of indicators depends on the quality of data and data sources used to generate them. The completeness of the description of each quality dimension of data and data sources can be indicative of the quality of the computed indicator, can help to harmonize data from different sources and improve their comparability.

V. CASE/EVENT/CONDITION AND INDICATOR DEFINITIONS

The *case definition* should consider the definition of the event of interest and the data necessary to determine the case/event/condition [for instance, symptoms, duration, electrocardiogram (ECG), autopsy are all elements used to validate a coronary event].
The case definition should also include a description of the following characteristics: variable type (e.g., numeric or string) and format (e.g., words, numbers, dates, times, percentages); mandatory fields, which can influence the proportion of missing data; and allowed values (range of allowed values or categories). A case definition depends on the objectives pursued by the specific study.
Case definition is a crucial element in the process adopted to build an indicator from data and data sources, because the epidemiological case definition often differs from country to country, or different time periods, or the clinical definition of the disease.

The *indicator* is a statistical synthesis of data according to defined rules, laws, and functions. The definition of the indicator should consider:

1) target population of interest (e.g., age, sex, other characteristics),
2) area under surveillance (e.g., geographical area, population size),
3) period under consideration.

As an example, Table 1 shows different case definitions for the "prevalence of obesity" *indicator* from different data sources.

Table 1. Prevalence of obesity from different data sources

Project	Case definition	Target population	Geographical area	Period	Obesity Prevalence %	
					Men	Women
EHIS(a)	ECHI	> 18 years	Europe	2006-2009	15,5	16,1
EHES (b)	EHES	25-64 years	Europe	2008-2012	18,6	18,1
Italian HES (b)	EHES	35-74 years	Italy, 20 regions	2008-2012	24.5	24.9
HIS -PASSI (c)	self-reported weight , height	18-69 years	Italy	2012-2015	11,1	9,8
HIS- ISTAT (a)	ECHI	> 18 years	Italy	2015	10,8	9
GPs (d)	measured weight, height	35-74 years	Italy	2008	23	23

Legend of table1:
EHIS=European health interview survey [13];
EHES=European health examination survey [14];
Italian HES= national health examination survey [15];
HIS-PASSI=health interview survey "Progressi delle Aziende Sanitarie per la Salute in Italia"-Italy [16];
HIS-ISTAT=health interview survey-Italian National Statistical Institute [17];
GPs=General Practitioners database [18].

(a) ECHI indicator: Proportion of adult persons (18+) who are obese, i.e. whose body mass index (BMI) is ≥ 30 kg/m^2. Calculation: BMI is defined as the individual's body weight (in kilograms) divided by the square of own height (in meters).
Case definition adopted by ECHI: Weight and height derived from European Health Interview Survey (EHIS) questions BMI01: How tall are you? (cm), and BMI02: How much do you weight without clothes and shoes? (kg). EHIS data are not age standardised [13].

(b) EHES indicator: Proportion of adult persons (25-64 years) who are obese, whose BMI is ≥ 30 kg/m^2. Calculation: BMI is defined as the body weight (in kilograms) divided by the square of own height (in meters) measured by scale and wall height ruler following standardised procedures and methods.
Case definition adopted by EHES: Weight and height derived from European Health Examination Surveys (EHES) [14,15].

(c) HIS-PASSI indicator: Proportion of adult persons (18-64 years) who are obese, whose BMI is ≥ 30 kg/m^2 [16].
Case definition adopted by HIS-PASSI: Self-reported weight and height derived by the questions submitted by telephone: Can you tell me your height without shoes? Can you tell me your weight without shoes and in underwear?

(d) GPs survey indicator: Proportion of adult patients (35-74 years) who are obese, whose BMI is ≥ 30.0 kg/m^2 or waist circumference is > 88 cm in women and > 102 cm in men [18].
Case definition adopted in GPs survey: weight, height, and waist circumference were measured using international standardised procedures and methods (e.g., standard electronic scale, wall height ruler) with the person in underwear [18].

VI. DATA AND DATA SOURCES QUALITY DIMENSIONS

Quality dimensions of data and data sources are listed below:

A. Relevance
B. Accuracy (validity, completeness, consistency)
C. Timeliness and punctuality
D. Accessibility and clarity
E. Comparability
F. Coherence

All the dimensions that influence the data quality are interrelated and influence each other, therefore partial overlaps are possible in the separate description of each dimension.
All quality dimensions are relevant for the use of primary data (e.g., ad hoc data collection by HIS or HES) or secondary data (data already collected, such as administrative data or data collected for different purposes). Dimensions can be different from one study to another, because different subdivisions or grouping of quality dimensions can be adopted. In this report, we refer to the quality list used by Eurostat [10].
The different dimensions of quality are described in detail here below, and examples are given.

A. RELEVANCE

Data should be relevant to the purposes for which they shall be used, and should respond to potential users' needs. This will entail a periodic review of requirements to reflect changing needs. Statistics that no longer have any interest for the set objectives should be discontinued and abandoned. Statistics about population health and health care performance are important if they significantly contribute to assess morbidity/mortality, disease-related quality of life, functional limitations (e.g. handicaps, limits to social participation), quality of care at practice or hospital level; if they are associated to a high utilization rate, or support the planning of health systems and economic resources.

Before starting a study, investigators should ask themselves important questions such as who will be the study's potential users, and which needs and methods the study shall consider.
About the users: 1) who and how many are the users? 2) How important is each of them?
About the needs: 1) what are the needs raised by the users and that the study shall meet?
About the method: 1) to which extent are the needs met? [10]

B. ACCURACY (validity, reliability, precision, completeness, consistency)

Accuracy indicates the closeness of the estimated value to the true value. Accuracy is a multi-faced dimension of quality, as it includes different aspects that are in some cases interrelated with each other. Validity, precision and reliability represent some aspects that contribute to accuracy, as well as completeness and consistency.

B1 Validity

Validity may be applied to a method or to an instrument; it indicates the extent to which that method or instrument measures or performs what it has been designed to measure/perform. During data collection measurements, errors may occur; they influence the validity because they generate values different from the true ones. For example, a laboratory method is valid when the values obtained are within an established range. Errors can be due to instruments (device or questionnaire), to respondents (giving a wrong answer consciously or unconsciously), to researchers/technicians (laboratory operators using wrong methodologies, interviewers who influence the answer). Instrument and researcher/technician errors can be assessed by repeating the measurement with a different instrument, or laboratory test, or interview performed by a different person. It is more difficult to assess respondent's errors, as this requires different sources for the same respondent. Data inconsistencies can be detected when editing data; data inconsistencies suggest the presence of errors. The proportion of records that fail each edit is an indication of data collection and data processing quality.

B1.1 Assessment of validity: agreement

Validity can be assessed also by measuring data consistency with the gold standard, i.e. the agreement of hospital discharge records (HDR) with medical records. Sensitivity (proportion of true positives that are correctly identified) and

specificity (proportion of true negatives that are correctly identified) are the two elements upon which accuracy can be estimated.

B1.2 Assessment of validity: missing information

The registered cases whose variable values are unknown create problems with data collection: this missing information is due to inadequate case histories, investigation or ambiguity in the medical record. If the quantity of missing data is huge, this may influence the results of the study. For example, incident rates of coronary events will be underestimated if a significant proportion of registered events appear in the "insufficient data" category rather than in the "definite" or "possible" or "probable" category. Missing information concerns proportion of missing data can be categorised as: 1) Missing Completely At Random, 2) Missing At Random, 3) Missing Not At Random.
Data are *Missing Completely At Random (MCAR)* when the probability of missing data on a variable is unrelated to other measured variables and to the values of the variable itself. Missingness of data is completely unsystematic, i.e. the probability that an observation is missing is not related to any other patient/people characteristic. For instance, an example of a MCAR mechanism could be a laboratory sample that was lost, and therefore 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 *Missing At Random (MAR)* when their 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 if 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 the proportion of cases with missing data in which the missing data tend to be greater amongst elderly population.
Data are *Missing Not At Random (MNAR)* when there is a relationship between the propensity of a value, or a range of values, to be missing and a specific value, or ranges of values. An example of a MNAR mechanism is the collection of information on annual income. Typically, people earning higher incomes may be less willing to reveal it, so the resulting observation is missing. In this case, missing values of annual income do not depend on other observed variables, but depend on a characteristic (higher income) of the incomplete variable itself.

Population based Registries: the lack of accuracy (at recording) can derive from missing basic variables such as date of event, site of cancer, sex, residence, date of birth; in the case of MAR, this occurs when the date of the event is missing for non-fatal events more than for fatal events; in the case of MNAR, this occurs when missing dates of the event are more frequent in specific seasons of the year than in other periods or the overall year, or when a low sensitive or low specific

instrument is used (e.g. signs or symptoms instead of nuclear magnetic resonance/computed axial tomography in stroke diagnosis).

*Health interview/Health examination survey: l*ack of accuracy can derive from the use of a low sensitive or low specific instrument, or from non-adherence to standardised protocols due to the person's unawareness of own condition, or the person's unwillingness to share information. Deficits in instruments or in laboratory performance may lead to a systematic bias and thus to a lack in accuracy and a low validity.

B2 Reliability and precision

Reliability is assured when the repetition of a method gives the same results in the same conditions. Precision is a manifestation of reliability, as it indicates how close the measured values are to each other. Therefore, a measure is considered as reliable when it is repeatedly applied to the same population and provides the same result in a high proportion of times.

B2.1 Assessment of reliability

Two procedures can be applied to assess reliability: a test-retest procedure (measuring is performed twice on the same object, and the agreement between the results quantifies the reliability); inter-rate procedure (different evaluators are used in independent measurements of a same object and the agreement between the results quantifies the reliability).

B2.2 Assessment of precision

Precision is the agreement among data collectors; it can be evaluated by re-extracting data from the same source (e.g., at a later time or by someone else), or from two or more different sources (e.g., coronary events from registry and acute myocardial infarctions from hospital discharge registry) and discrepancies are assessed by comparing the various data sources.
In HES, in order to increase precision, the following methodologies can be used: regular monitoring of measurers performance; regular reviewing of instruments; periodic assessment of laboratory performance; data input with variable control for ranges.

B3 Completeness

Different definitions exist for completeness of data sources. We adopt the following: completeness describes the degree to which values are present in a data collection [19]. Completeness can be referred to the event or to the information.
In registries, completeness of the events can be assessed by the 'coverage rate' and describes the extent to which all the expected events are registered. It is measured as percentage of registered events divided by the expected ones. The best condition is 100%. A control of the coverage allows to identify missing or duplicate events. A method to control event completeness is the record linkage with other sources of information (HDRs linked for non-fatal events and mortality for fatal events). In a population-based registry of coronary events, non-fatal events occurred out of the surveillance area are rarely recorded; in that case, to ensure completeness, General Practitioners (GP) databases may be used to catch the event after the acute phase.
Completeness can be also referred to the information recorded for case/event definition and validation. For example, in the case of coronary events, the International Classification of Diseases (ICD) codes reported in HDR and in mortality and date of event onset are necessary to establish the event; according to MONICA criteria [20], symptoms, enzymes, ECG, autopsy (in case of fatal events) are all necessary elements to validate a suspected event.
Case-finding may be problematic in multimorbidity patients, for example in elderly people, in which the presence of multiple pathologies can make a single diagnosis by hospital admission/discharge record more difficult to classify.
In HIS and HES, completeness can be invalidated by coverage errors caused by divergences between the target population and the frame population, defined as "the set of population units which can be actually accessed and the survey data that refer to this population". Ideally, the frame population coincides with the target population; this situation is difficult to obtain because the frame population is smaller than the target population.

The coverage errors can be distinguished in:

1) undercoverage: persons who are not accessible by the frame, for example persons resident in a given area, but temporarily out of the area;
2) overcoverage: persons who are accessible by the frame but who do not belong to the frame (e.g., inclusion of dead people);
3) multiple listings: persons who are present more than once in the frame (e.g persons with two or more telephones);
4) incorrect auxiliary information: persons with wrong information.

Coverage errors can lead to bias and underestimation of the variance. Overcoverage, multiple listings, and incorrect auxiliary information can be avoided

checking the information about each unit of population. Undercoverage is more difficult to detect and specialised frame quality reviews are necessary to discover them.

The main methods to evaluate completeness are qualitative and quantitative.

B3.1 Assessment of completeness: qualitative methods

Qualitative methods estimate the degree of completeness by comparing study data to other data sources, or over time.

- Geographic data analysis: it includes the comparison of event rates in one population with rates observed in other populations in which similar rates are expected. Deviations from regional standards may reflect specific local variations in risk factor prevalence or in the frequency of screenings carried out for some high risk conditions; anyway, systematic discrepancies (across several sites) provide evidence of possible under-registration (or over registration due to inclusion of duplicate records);

- Historical verification: it includes the comparison of event rates with those observed in the same population in a different period of time;

- Ratio between mortality and incidence rate: it is an example of independent case ascertainment; when the ratio between mortality and incidence is greater than expected, this may lead to a suspicion of incompleteness (incident cases missed by the registry);

- Number of sources/notifications per case: using as many sources as possible reduces the possibility of unreported cases and increases the registry data completeness. An efficient record linkage is essential.

B3.2 Assessment of completeness: quantitative methods

For registries and for HIS/HES, quantitative methods assess the extent to which all eligible cases have been registered.

- Independent case ascertainment: the sources of information are rechecked to detect possible cases missed during the registration; one or more independent sources of cases are used and databases are compared (cases recruited in international clinical follow-up studies, patients enrolled into a multicentre clinical trial, databases of GPs, patients enrolled in cohort studies, patients

enrolled in multihospital case-control studies, patients enrolled in community screening).

- Capture-recapture method: this method was originally developed as a method to estimate an animal's population size. The procedure can be described as follows: in a defined area, as many animals as possible are captured, tagged and released (capture stage); at a later time, this procedure is repeated (recapture stage). The numbers of animals in each sample and the numbers of animals common to both samples (recaptured) are used to estimate the overall numbers in the total population (assuming that capture and recapture are independent); capture and recapture methods can be used in epidemiology to estimate the extent of complete ascertainment of disease registries [21]. Typical applications include estimating the number of people needing particular services (i.e., services for children with learning disabilities, services for medically frail elderly people living in the community), or experiencing particular conditions (i.e., illegal drug addicts, people infected with HIV, etc.). For example, to implement a register of children with Type 1 diabetes, children are identified from hospital admission records, from GPs (family doctors) databases, and from the records of the local Diabetes Association. None of these sources have a complete list, but by putting them together it is possible to obtain and to estimate how many children are identified in total, how many children with Type 1 diabetes are living in the community, and to which extent each data source can provide a complete ascertainment.

Examples of completeness measures for registries and HIS/HES:

- Population-based registries: cancer registries provide the percentage of cases without microscopic diagnosis (completeness of information); cardiovascular diseases (CVD) registries provide the number of fatal cases with insufficient data (completeness of information); number of patients admitted in hospital out of the surveillance area (completeness of cases);
- Health interview/Health examination surveys: assess the proportion of persons examined in the eligible population (participation rate); incompleteness of collected information or exams (percentage of missing data for each variable).

B4 Consistency

This dimension describes data plausibility. Data consistency can be checked in a variable (internal consistency), between different variables, or at two or more points in time (historical consistency). Most quality checks performed for single variables concern format and allowed values, but can be also more specific and

refer to dates and classifications. Consistency can be affected by processing errors that can occur between the time of data collection and the beginning of the statistical analysis. Processing errors can be present in each single step: coding, data entry, data editing, imputation, etc.
To evaluate the impact of the error on final statistics, data should be re-coded/re-entered in the computer, or re-imputed, and errors corrected. A correction can be performed by assigning multiple imputed values to wrong or missing data.
Sometimes it is difficult to distinguish between processing and measurement errors, as they can overlap.

Example of internal consistency (single variable):

- Dates: if this variable contains day/month/year all together, it is necessary to check each one of them. If the month is January, March, May, etc., the range values for the days must be (1-31). If the month is April, June, September, the range values for the days must be (1-30). For example, if 31/09/2016 is found, there is a mistake, because this date doesn't exist;
- Age at diagnosis: it is usually calculated as year of diagnosis minus year of birth; this procedure rounds up age by 6 months. The age range values must be positive and between 0 and 100.

Example of consistency between variables:

- Population–based registries
  - consistency between dates (birth, diagnosis, death, autopsy): date of birth should be $\leq$ date of diagnosis; date of birth should be <= date of death; date of death should be $\leq$ date of autopsy;
  - consistency between age and sex: some diseases/conditions occur almost exclusively in specific age groups (children, adult, elderly); others only in men or in women (i.e., prostate hypertrophy, menopause).
- Health interview/Health examination Survey
  - consistency concerning a condition: the 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 between ages: the computed age at diagnosis should be consistent with the registered age.

C. TIMELINESS AND PUNCTUALITY

Timeliness relates to the rapidity of data collection, the processing and reporting of reliable and complete data, and the length of time between data collection and dissemination of results. Speedy access to data and indicators is a priority and a clear benefit to health providers, investigators or policy makers. Early provision also enhances the reputation of the registry and the data source; however there is a trade-off between timely data and the extent to which they are complete and accurate. Punctuality refers to the lag time between the scheduled date, established in a calendar (reference date), and the actual delivery dates. Usually data collections have predefined time intervals (year) that should consider a balance between timeliness and completeness. Timeliness and punctuality influence the frequency of released statistics that depend on the time needed to plan and perform the survey or to collect complete data for a registry, or the time to perform the quality control of data, statistical analyses and interpretation of results.

In population-based registries, timeliness refers to the length of time between the event occurrence and the dissemination of results; e.g., registries are constantly updating their databases, as they receive new reports, but some notifications, especially those from deaths of residents out of the area, arrive long after the case was diagnosed. Timeliness in the publishing of results can be related to completeness: registries may have a tendency to delay the dissemination of their results in order to achieve better completeness. The availability of quantitative methods that allow to estimate completeness at a given stage of the registration process could help registries to decide in a more rational way when data can be considered ready for publication.

In *surveys*, timeliness refers to the length of time elapsing between data collection and dissemination of results (e.g. prevalence and mean risk factor distribution). Usually, in surveys, timeliness and punctuality may be affected by delays in data collection and data quality control. If data are required for policy makers, it is important to make them quickly available to all interested stakeholders.
If data must be used for research purposes, - such as, for instance, serial analysis- then it is important to make them quickly available to all the interested researchers, so to avoid their use by a single researcher rather then by more researchers in parallel.
A delay of some years has usually less importance in the etiological study implemented to support policy-makers in planning preventive actions, than in surveys aiming at evaluating efficacy of preventive or health care performance.

D. ACCESSIBILITY AND CLARITY

Accessibility refers to users' easiness of access to statistics, as well as the suitability of the form or media through which information can be accessed. Therefore accessibility relates to different aspects of data dissemination, such as the distribution channel, the marketing conditions (i.e., copyright, etc.), ordering and delivery procedures, pricing policies, availability of micro or macro data, formats (i.e., paper, files, CD-ROM, Internet), etc.

Clarity refers to the presentation of statistics in an understandable and clear manner. Clarity presupposes that statistics are accompanied by textual information and explanations, graphs, figures and other illustrations, offered by the data provider to assist users (e.g. provision of easy access to meta-data). Usually, documents tend to be written in a language and in a communication way understandable only to experts. A rigorous scientific communication is essential to share results among scientists and to provide evidence for guidelines, national and international reports and publications, as well as support to policy makers in planning prevention and care programmes. Efforts should be also pursued to make results user-friendly also to other stakeholders, i.e. health operators, journalists, patient associations and interested persons. If data cannot be accessible, or the associated metadata are not understandable, even the most accurate and coherent data will have little value.

Accessibility and clarity are the most neglected quality dimensions, though the validity of a registry or survey is influenced by these two dimensions that, if well developed, allow a more detailed analysis and a broader dissemination of data to relevant audiences.

E. COMPARABILITY

Comparability is the extent to which the differences shown by statistics coming from several geographical areas or non-geographical domains, or collected over different periods of time, can be attributed to differences between the true values of the statistics [17]. Data comparability is a crucial aspect to any reliable conclusion and benchmarking between countries/regions, and over time. The lack of comparability can be related to factors such as: 1) use of different definitions, 2) use of different procedures or measuring tools, 3) use of different standardisation or computation methods (for instance, to get life expectancy measures).

Comparability can be ensured by applying proper methods for the standardisation and harmonization of collected data. These methods have to be clearly defined and described in detail, such as in the HA4 report: "Standardisation methods for the collection of health information".

For disease-specific data, the basic requirement is the standardisation of the case definition. For example, in the recommendations provided by the WHO registry for Acute Myocardial Infarction (AMI 1976), the WHO-MONICA (MONItoring trends and

determinants in CArdiovascular diseases [20]) Project (1980-2000), and the EUROCISS (European Cardiovascular Indicators Surveillance Set [6,22]), three different definitions of events were adopted:

- in the WHO registry for AMI, only the anamnestic history was collected and the ECG was not used for the identification of the event.
- in the WHO MONICA Project, where, for the identification of events, the disease specific data collection was based on hot and cold pursuit; enzymes, evolution of ECG, symptoms, and autopsy were used as diagnostic criteria for the validation of the event [20].
- in the EUROCISS Project, a step-wise procedure was recommended, based on: i) standardised data collection; ii) appropriate record linkage between hospital and mortality records or other sources of information (e.g., GPs, drug dispensing registers); iii) a selection of a random sample of suspected events and their validation through the MONICA method and the new diagnostic criteria of the European Society of Cardiology/American College of Cardiology (ESC/ACC) assessing the positive predictive values (PPVs) of the ICD codes reported in death certificates and in hospital discharge records, and using PPV to estimate the number of current events [22,6].

A precise knowledge of current and historical registration procedures, methods and definitions is of great importance also in the analysis of the geographical and temporal variation. Geographical comparability refers to the comparison of similar surveys that analyse the same phenomena, but involve the population of different geographical areas, or are conducted by different organisations and/or include different time periods.
As concerns comparability over time, the data collected in a specific reference period cannot be fully comparable with those collected in subsequent periods, if changes occurred, and, consequently, a break in the time series is introduced. Changes in references, concepts or measurement processes should be documented and their impact should be assessed. An example is reported for the AMI Registry: the availability of more sensitive tests provided a better assessment of the severity of non-fatal disease and therefore modified the results obtained for the incidence of the disease over time; today more treatments are available for the acute phase, and therefore the detection of mild events (Acute Coronary Syndromes defined by the troponin test) can be carried out better than in the past.

To assure comparability, particular attention should be given to:

- definition of area under surveillance and target population;
- the case definition used (e.g., ECHIM indicators, international guide-lines), rules for coding events (e.g., sources, algorithms, version of international classification of diseases), definition of multiple events in a same individual (e.g., for coronary or cerebrovascular recurrent events, the threshold of the

28th day is internationally used), and the date when the disease becomes an event -'onset of the event'- (first date of admission at hospital, first prescription, diagnosis by the GP, until the date of death);

- detection of asymptomatic events: incidental detection of event (e.g., asymptomatic myocardial infarction or cancer can be detected during HES or a screening programme);
- autopsy performed with or without consent: in some countries, autopsies are frequently performed without consent, for medical, scientific or educational reasons; in these countries or regions, the number of events may be higher.

In order to improve data comparability, extensive descriptions of the methods used to produce data, data sources, and methods used to compute indicators should be periodically published and updated. Description of methods should be published on web-sites, but also in the reports drafted by the research institute and in national and international journals, as well as in project protocols, in order to avoid loss of methodologies adopted over time (web pages may change over time).
Statistics from different domains can be compared taking into account that different concepts could be used (definition of characteristics, reference period, etc). All differences should be reported and their effects evaluated.
Measures and remarks on data comparability should be accompanied by the list of health indicators; e.g., remarks on comparability between countries, over time and with national data, are provided for indicators in the ECHI shortlist, where applicable. Data comparability is a crucial to get reliable conclusions and perform a benchmarking between countries/regions and periods.

F. COHERENCE

In registries and surveys results, coherence refers to their possibility to be reliably combined in different ways and for various uses and purposes. The coherence level of statistical information is proportional to the possibility of combining this information and other statistical information in an analytical framework and over time [23].
Coherence is promoted by the use of standard concepts, classifications and target populations, as well as the use of a common methodology across registries and surveys. Coherence does not necessarily imply full numerical consistency. Coherence reflects the degree to which data and information from a single study are brought together with other data and information, and how they are logically connected and completed. Fully coherent data are consistent internally, over time and across results, products and programmes.

Coherence can be assessed for different areas:

a) coherence between temporary and final results, to establish whether the difference between temporary and final statistics is really significant;

b) coherence between annual and short term statistics;
c) coherence between statistics in the same domain: when a group of statistics, possibly of a different type, measures the same phenomenon with different approaches;
d) coherence among survey data compared at national level.

Coherent statistics validate the respective data they contain and can be validly used and combined.

VII. QUALITY REPORT AND QUALITY INDICATORS

In data assessment, a clear picture of data quality and the methods used to assess quality are needed; the definitions and dimensions discussed in *Section VI* are preconditions. A report on data quality is crucial, and should describe data characteristics according to quality components, as well as present these characteristics according to quality requirements.
The quality report summarises the most important information on quality. The measurable aspects of quality can be characterised by indicators, and the information contained in the report helps to understand the limitation of a given result. Self-assessments, audits and peer reviews are based on information obtained from quality indicators and reports, process variables, and user surveys. Furthermore, they sometimes might use specifically designed checklists in order to present the information needed in a more structured and accessible way. According to the Eurostat Standard quality report for Labour Force Survey [24], a detailed quality report should include:

- Administrative information
  - The name, the reference period and the periodicity of registries or surveys
- General Description
  - The design and methods used for the survey
  - A description of the methods used during the survey process (classification, sampling design, data collection process, etc.)
- Relevance
  - A description and the classification of the users
  - A description of the variety of the users' needs
  - Main results regarding the satisfaction of users
  - *Indicators:*
    - user satisfaction index
    - rate of available indicators

- Accuracy
  - Sampling errors
    - Order or magnitude (or at least sign) of the bias of the main variables
    - Methodologies applied for variance estimation
    - *Indicators:*
      - estimated coefficients of variation (CV) for the quantitative variables
      - imputation rate and ratio
  - Coverage errors
    - Type and size of coverage errors
    - Information about the frame: reference period, updating actions, quality review actions
    - *Indicators:*
      - over-coverage and misclassification rates
      - geographical under-coverage ratio
  - Measurement errors
    - The measurement errors identified and their extent
    - Indications about the causes of measurement errors
  - Processing errors
    - A summary of the processing undergone by data between their collection and the production of statistics
    - Identified processing errors and their extent
    - *Indicator:*
      - average size of revisions
  - Non-response errors
    - Non-response; unit and item non-response rates for the main variables, both unweighted and weighted
    - Imputation methods used (if any)
    - Indications about the causes of non-response
- Timeliness and Punctuality
  - *Indicators:*
    - the average timeliness of data (time lag between the end of the reference period and the date of first results; time lag between the end of reference period and the date of the final results)
    - Data frequency and average data freshness
    - Punctuality of time schedule of effective publication

- The reasons for late delivery

- Accessibility and Clarity
  - A summary description of access conditions to data: media, support, marketing conditions, existing service-level agreement, etc.
  - A summary description of the information accompanying the statistics (documentation, explanation, etc.)
  - *Indicators:*
    - Number of publications disseminated and/ or sold
    - Number of accesses to databases
    - Rate of completeness of metadata information for released statistics

- Comparability
  - The reference period of the survey where the break occurred
  - The differences in concepts and methods of measurement before and after the break
  - *Indicators:*
    - Length of comparable time-series
    - Number of comparable time-series
    - Rate of differences in concepts and measurement from European norms
    - Asymmetries for statistics mirror flows

- Coherence
  - Coherence of statistics in same domain: summaries of the mirror statistics
  - Coherence with National Accounts: a summary of the comparison
  - *Indicators:*
    - Rate of statistics that satisfy the requirements for the main secondary use

- Cost and Burden
  - Cost supported by National Statistical Institute (NSI)
  - Response burden.

Quality indicators are specific and measurable elements of statistical practice that can be used to characterise the quality of health indicators.
Quality indicators specify the quality components of a product and make its description more informative and transparent. Users can assess the quality of the different surveys or the same data in different periods by using these quality indicators.
Some quality indicators should be produced for each output, in line with the frequency of statistics or publication (for example, standard errors should be calculated for each new estimate). However, some quality indicators should be

produced only once and for longer periods, and should only be rewritten when major changes occur (e.g. lag time between the end of the reference period and the date of first results). The frequency of the indicators calculation depends on the purpose of quality indicators (e.g., monitoring the quality over time) or on the survey or publication frequency.
This quality measurement should be part of all statistical processes and should not be a separate activity carried out after the statistics are produced or when users need it. This should be done not only for the sake of cost and time efficiency, but also for quality improvement purposes.
The way in which quality reports are disseminated must be integrated in the dissemination policy of registries or surveys. Registry and survey experts should work together in developing quality reports and indicators. In the short-term, a detailed quality report allows registry operators to improve data collection. Quality measurements should be done at an acceptable level in terms of expenses and time consumption. To do so, it is better to start with those indicators and components that are more important. Indicators may be misleading, or they may focus only on a part of the phenomena. For example, as concerns accuracy, more attention is generally given to sampling errors than non-sampling errors; however, the latter may have a dominant role, even though it is difficult to measure them. At a later stage, the use of this simplified quality report can be extended to all dimensions. Finally, the level of detail and content of the report has to be improved to meet the specific needs of its users.
Quality reports need time and efforts; they are efficient only if used in an appropriate way: their level of detail, structure and form must be suitable for the targeted users (not too long, easy to understand, simply and clear). When quality indicators are used to inform users on the quality of statistics, it is recommended to include qualitative statements that can help to interpret quality information and summarise the main effects on the usability of the statistics.
The use of a same quality report template, with a limited set of standard indicators for different products, will support transparency. Quality indicators and quality reports include the most important information on quality; this is the reason why they are used very often as the basic documents for self-assessments or audits.

VIII. QUALITY METHODS DEVELOPED IN EU PROJECTS: IMPLICATIONS AND LIMITATIONS

Some of the EU (funded) projects on health information are based on ad *hoc surveys* (primary use of data), or on integrated health information systems (secondary use of routine data), or on linkage of different sources of information and ad hoc surveys (secondary and primary use of data), and therefore a short description of the projects and a literature review is presented, together with

their possible implications, limits and recommendations according to quality dimensions.

Other projects did not collect data, but simply recommended procedures and methods to collect high quality and comparable data, and methods to improve and maintain quality over time.

Table 2 describes the information collected by the questionnaire sent to the principal investigators within BRIDGE Health, regarding data collection tool, quality methods described in the manual of operations, and references.

Table 2. Quality methods in the EU Projects. Sources: BRIDGE Health questionnaire and references

EU Project	**Data collection tool**	**Primary or secondary use of data**	**Quality methods adopted (manual operations, websites literature)**	**References**
EHLEIS	HIS, Registries	Secondary	*Completeness* of information and events; *Data consistency/coherence;* *Internal validity;* *Personnel training.*	
	Different data sources	Secondary	*Completeness* of information and events; *Data consistency/coherence;* *Internal validity;* *External validity.*	
EHES	Population based health examination survey	Primary	*Completeness* of information and events; *Personnel training and* site visits.	Tolonen H [25] Tolonen H [26]
EHIS	Health Interview Survey	Primary	*Manual for planning and implementation, including conceptual guidelines, a model questionnaire, a translation protocol, interview instructions and statistical survey guidelines. Quality report templates provided for quality reporting.*	Commission Regulation (EU) [27] Eurostat [28]
Injury Surveillance-JA on Monitoring Injuries in Europe	Routine care data and some disease registries	Primary and Secondary	*Completeness* of information and events; *External validity;* *Personnel training.*	Nwaru BI [29]
Injury surveillance platform/Euro Safe	Routine/ad ministrative data sources	Secondary	*Completeness* of information and events; *Internal validity;* *External validity.*	EuroSafe [30]
EUROCISS	Register data on AMI and Stroke	Primary	*Representativeness;* *Standardised procedures and methods;*	Madsen [22] Giampaoli [31]

			Completeness of information and events; *Internal validity;* *External validity;* *Personnel training.*	
	Health examination survey	Primary	*Representativeness;* *Completeness* of information and events; *Internal validity;* *External validity;* *Personnel training.*	Primatesta [32]
COPHES/DEMO COPHES OBELIX/ENRIECO FLEHS	HIS; Examinations surveys (hair and morning urine)	Primary	*Completeness* of information; *Internal validity;* *Personnel training.*	Becker [33] Casteleyen [34] Esteban [35] Exley [36] Fiddicke [37] Schindler [38]
EURO-PERISTAT	Different data sources	Secondary	*Completeness* of information and events; *Internal validity;* *External validity;* *Personnel training.*	Gissler M [39] Euro-Peristat [40]
ECHO-European Collaboration for Health Care Optimization	Routine/administrative data	Secondary	*Completeness* of information and events; *Internal validity;* *External validity.*	ECHO [41-43]
EUROHOPE	Routine/administrative data	Secondary	*Completeness* of information and events; *Internal validity;* *External validity.*	Häkkinen U [44] EuroHOPE [45]
ECHI-1, ECHI-2, ECHIM, JA-ECHIM	Different data sources	Secondary	*Completeness* of information and events; *Internal validity;* *External validity;* *Personnel training.*	ECHI [46]
EuroREACH	Different data sources	Secondary	*Completeness* of information; *Internal validity.*	EuroREACH [47]
EUBIROD Network	Different data sources	Secondary	*Completeness* of information and events; *Internal validity;* *External validity.*	Carinci F [48] Cunningham SG [49]

A. EURO-PERISTAT

The EURO-PERISTAT Project monitors health and care of mothers and babies during pregnancy, delivery and post-partum period [39]. Thirty-one countries currently participate in the project, including all current EU member states, Iceland, Norway, and Switzerland. Bulgaria and Croatia joined the network in 2015.

Indicators are grouped in four categories: neonatal health; maternal health; population characteristics or risk factors; healthcare services. Each category contains core indicators, recommended indicators, and further indicators in development (Table 3). The indicator definitions and data are available from the Euro-Peristat website (www.europeristat.com)

Table 3. EURO-PERISTAT, list of indicators, updated 2012

Category	**Core**	**Recommended**	**Further development**
Neonatal health	C1-Fetal mortality rate by gestational age, birth weight, plurality C2-Neonatalmortality rate by gestational age, birthweight, plurality C3-Infant mortality rate by gestational age, birthweight, plurality C4-Birthweight distribution by vital status, gestational age, plurality C5-Distribution of gestational age by vital status, plurality	R1-Prevalence of selected congenital anomalies R2-Distribution of APGAR score at 5 minutes R3-Fetal and neonatal deaths due to congenital anomalies R4-Prevalence of cerebral palsy	F1-Severe neonatal morbidity among high risk infants F2-Prevalence of neonatal encephalopathy F3-Causes of foetal and neonatal death other than CA
Maternal Health	C6-Maternal mortality ratio by mater age	R5-Maternal mortality ratio by cause of death R6-Prevalence of severe maternal morbidity R7-Prevalence of tears to the perineum	
Population characteristics or risk factors	C7-Multiple birth rate by number of foetuses C8- Distribution of maternal age C9- Distribution of parity	R8-Percentage of women who smoke during pregnancy R9- Distribution of mothers' education R10-Distribution of households' occupational	

		classification R11-Distribution of mothers' country of origin R12-Distribution of mothers' body mass index (BMI)	
Health care services	C10-Mode of delivery by parity, plurality, presentation (of foetus), previous caesarean section	R13-Percentage of all pregnancies following sub fertility treatment R14-Distribution of timing of 1st natal visit R15-Distribution of births by mode of onset of labour R16-Distribution of place of birth by volume of deliveries R17-Percentage of very preterm infants delivered in units without a NICU R18-Episiotomy rate R19-Birthswithout obstetric intervention R20-Percentage of infants breast-fed at birth	F4-Neonatal screening Policies F5- Content of antenatal care

Implications

This Project demonstrates the feasibility and value of indicators to monitor perinatal health at a European level, as data on these indicators have been collected for three publications (for the years 2000, 2004 and 2010). However, the results of this project also illustrate that continuing international collaboration is needed to improve the consistency of definitions and to prioritise the development of data collection methods for many perinatal health indicators. While all indicators can be provided by at least a few countries, no country can provide the full set of indicators. The core indicators are more widely available.

Quality is ensured by the use of common pre-established definitions, the collection of data in terms of numbers of births (as opposed to percents already calculated), so that numbers can be cross-checked across indicators (numbers of live births, stillbirths, etc) and so that the coordination team can be sure that percentages are calculated in the same manner. The numbers of missing data are also collected for each indicator and reported in the tables.

When countries cannot provide the Euro-Peristat data using the agreed definition, they are requested to provide data using their national definition and describe their definition.

For each data source, information is requested on population and coverage, and on whether evaluations of its quality (coverage, completeness, external validity) have been undertaken.

Finally, quality is maintained by using data to produce scientific articles. While analysing the specific indicators and comparing them with others, as well as with the scientific literature, outliers are identified and discussed with scientific representatives and other experts. This allows the group to get a better sense of the strengths and limitations of our data. More than 50 articles have been published in peer reviewed journals using data from the Euro-Peristat project.

Limitations

The quality method problems in PERISTAT are related to case definition and coverage of data collection, which influence completeness of data collection.

Case definition: WHO criteria for stillbirth is foetus with a birth weight of 500 gr, or - if this is missing - a gestational age of 22 weeks. This legal limit for registration is not always respected by the various countries, who used a time period exceeding the 22 weeks (24 weeks Hungary, Portugal, UK, 25 weeks + 5 days Italy, Luxembourg, Spain, 28 weeks Greece, Sweden).

A standard definition of "stillbirth" is essential when international comparisons are made. For live births, most countries have no limits for weight and gestational age, while others have specific inclusion values.

However, Euro-Peristat collects the information required to create comparable indicators by using thresholds that can be applied in all countries [50,51].

Another discrepancy depends on the different criteria used by civil registration and health registration systems for the inclusion of non–residents. Civil registration is limited to permanent residents only, while health systems comprise all the events. This difference can influence data, where a large number of non-residents are present (refuges, immigrates, visitors, asylum seekers).

Recommendations

Given the large proportion of deaths before the 28th week (over one-third of all deaths), it is essential to improve information systems in Europe by developing common guidelines for recording births and deaths at 22 weeks.

A continued collection of the full set of Euro-Peristat indicators is a drive for countries to improve their national systems, so that they can produce the key indicators that are available in other countries and acquire a better knowledge of

the studied phenomena by comparing their perinatal health systems with the health systems of their neighbours.

B. EUROREACH PROJECT

EuroREACH is a project whose objective is to improve access to, and use of, healthcare data, and enhance cross-country comparisons of health system performance [52]. It has produced the "Handbook to access health care data for cross-country comparisons of efficiency and quality" and then a digitised form of the Handbook, the Health Data Navigator – HDN (www.healthdatanavigator.eu) [53] to facilitate the dissemination of the outputs to the research community. The EuroREACH appraisal criteria used by HDN were: governance, access to database, coverage, linkage, data quality, strengths and weaknesses. In terms of data quality, the following points are considered: entry errors (multiple common entries, redundancies that lead to incorrect/incomplete data), breaks (changes of standards that lead to incompatibility), consistency of terminology (differences in how data are described when collected).

The EuroREACH Case Study on diabetes care has pointed out a number of problems that can have an effect on the comparability of data from different health systems.

a) The source of data can be delivered by different systems, and this influences the comparison. A system with a centralized laboratory is more likely to provide laboratory data that allows for quality monitoring than a system in which laboratories contract with payers.
b) There are differences in coding systems between countries. Therefore a complete list of all codes that can identify a specific diagnosis should be available.
c) The type of hospital reimbursement can influence the level of detail with which diagnoses and procedures are coded. Prospective payment formula of hospital reimbursement generate data with more precise diagnosis and procedure coding than those using a per-day payment rate.

The EuroREACH case study for diabetes is an useful example to explore the practical problems faced when administrative data are used to compare the performance of chronic disease care in an international context.

Due to comparability and availability problems, only three countries (Finland, Estonia, Maccabi - Israel) participated in the study comparison of diabetes care at international level.

The study has revealed differences in the health care system/health care definition of diabetes, depending on severity or disease stage. Among the

participating countries of the EuroREACH diabetes study, Estonia did not have any disease Registry (except cancer) and provided data from the administrative health service Database of Estonia Health Insurance. No data linkage was necessary for the identification of the cohorts. Finland, on the contrary, has a system of medical registers and administrative database. Linkage was performed by a unique national personal identification number. However, it must be stressed that even if in some countries Registers for primary care (e.g. ICPC) exist since decades, the quality of diagnosis coding is not always satisfactory. Consequently, the definition of diabetic population requires the use of medication data from the Social Insurance Institution. For this specific study, a linkage-based research data base of Finnish diabetic patients was already available, hosted by the National Institute for Health and Welfare and its partners. Israel has a patient system covering visits to hospitals, physicians, prescription drug purchase, laboratory testing (with results), and imaging studies. For international data comparison, all databases were linked using the national identifier. The Bridge Health project, however, has no work package that can be expected to provide a full overview of European primary care registries. This limits the scope and validity of our conclusions.

Problems in data comparability can be tackled by using the Health Data Navigator instrument. These problems can be grouped in: 1) identification of data generating processes (data sources: administrative system or health record); 2) case definitions (use of primary classifications: data specifications defined nationally by the authorities are not similar); 3) comparison of cost and resource used; 4) stratification/risk adjustment [52,53].

C. EHES - European Health Examination Survey

The European Health Examination Survey is an initiative to set up a system of standardised, representative health examination surveys of the general adult population in the European countries. The core measurements are height, weight and waist circumference to measure body composition, blood pressure, total and HDL-cholesterol and fasting glucose. EHES includes also a self-reported questionnaire with important items needed to support objective measurements as well as health status and health perception. EHES is a survey with primary use of data collection, and provides comparable indicators on risk factors, chronic disease prevalence, physical performance, cognitive function, etc.

Implications

With a pilot study conducted in 14 countries, the Project demonstrated the feasibility of a survey conducted with standardised procedures and methods in samples of the general adult population in different countries [54,55,56].

Data quality is assured by the training and testing of the personnel involved in the fieldwork, through the application of international standardised procedures and methods, and with internal and external quality control. In this way, the study of time trends and geographical variation can be assured. Completeness is assessed by the participation rate, which depends on age, sex, season of survey, at home or at centre examination, time and day of the week, etc.. Audits and site visits are used to check quality and improve standardisation. An evaluation is also carried out for risk factors, risk conditions and prevalence benefits from use of primary data collection.

Limitations

Some countries have a long tradition in health examination surveys, therefore they prefer using methods or devices from their previous surveys to measure risk factors, and these methods/devices can occasionally differ from provided standardised recommendations. Changes in procedures and methods can affect the study of time trends of risk factors. A problem revealed by the project is the comparability of national HESs, due to differences in age groups and survey methods. Some of the measurements are sensitive to the protocols and devices used for the measurement. Blood pressure measurement is an example of this, as it is influenced by the participant's activity before measurement, posture during the measurement and the used device for the cut-off level (see example *Table 4* at page 50) [25,26].

Recommendations

To minimize the differences between the various national HES and optimize their comparability, a proper standardisation of the measurement protocols, personnel training and periodical internal and external audits are required.

D. EUROCISS - European Cardiovascular Indicators Surveillance Set

The aim of the EUROCISS project is to prioritise the aspects of cardiovascular diseases of major interest at European level, as well as provide a list of recommended indicators, sources of information, case definition and quality methods for monitoring AMI/Acute Coronary Syndrome, stroke and cardiovascular diseases surveys [22,31,32].

The project's main objective was to prepare the manuals of operations, which provided simple and comparable tools to support and stimulate the implementation of surveillance systems in those countries where they are not present, using administrative databases, such as mortality and HDRs, and validating random samples of suspected events. Starting from a minimum data set (MDS) and following a step-wise procedure, EUROCISS provided a standardised

model for an efficient implementation of a validated surveillance system at a reasonable cost. To set public health priorities and determine appropriate actions, a standardised definition of event is crucial, and indicators should be comprehensive, valid (sensitive and specific) and standardised, and they should meet quality criteria. The definition of the event must take into account both the ICD codes reported in (main or secondary) hospital discharge diagnoses and the (underlying or secondary) causes of death, as well as the duration of the event (28 days). This definition is of particular importance since myocardial infarction may occur more than once, and it is therefore necessary to consider both first and recurrent events. Inaccuracy increases with age. A unique person identification number (PIN) for each subject is a strong tool in linkage procedures between hospital discharge diagnoses and death certificate data; alternatively, multiple variables (e.g. name, date of birth, sex, residence) can be used for record linkage (deterministic and probabilistic).The high cost of registers limits their implementation at national level; therefore they should be established in representative areas of a country (regions, macro-areas, etc.).

Implications

Population-based registries are the best data source for cardiovascular surveillance as they include morbidity and in- and out-of-hospital mortality. They provide estimates of key indicators of population health, such as incidence/attack rates and case fatality. Incidence provides new events in the population at risk. Attack rates consider both first and recurrent events; case-fatality considers in-hospital mortality and sudden death, i.e. those serious cases that do not reach medical services. Incidence can be assessed when information on first event is available. If survival rates are available, also prevalence can be assessed.

These data are useful to study the time trends and geographical gradients of major cardiovascular diseases, and make it possible to assess reliable and comparable indicators to be used for research purposes; validation of diagnostic information is recommended in a random sample of a sufficient size of the identified events, with the estimation of sensitivity and specificity and PPV of the defined events; when the validation of a sample of events is not possible, and therefore PPV, specific of the area, cannot be estimated, the available PPV, taken from other registries or studies, can be applied to the identified current events in order to estimate the number of occurred events in the population.

Limitations

- Lack of registration of non-fatal events occurred outside the surveillance area or in nursing homes and clinics, and lack of registration of asymptomatic events. When hospital records are collected at national level, it is possible to include also those non-fatal events that occur out of the surveillance area;

- conduct of a propaedeutic but unavoidable activity before implementing record linkage, which consists in accurately checking and cleaning administrative databases of both mortality and HDRs from possible mistakes in identifying variables used for record linkage (family name, name, date of birth, place of birth, residency). This is necessary to avoid possible double counting of the same record or, on the contrary, to avoid that corresponding records do not link due to possible mistakes in corresponding identifying variables. These kinds of mistakes can heavily bias results, since they influence the identification of the first event, the dates of the first event and recurrent events, and consequently the number of events for the same subject and for the overall population included in the registry;

- obtaining administrative files for research purposes: mortality data files are usually available at the National Institute of Statistics, while hospital discharge data are available at the Ministry of Health. These data are anonymous and therefore do not allow record linkage. Nominal files of both mortality and hospital discharge records are available at the regional level or at the sanitary units. When data are combined, the missing events are mainly explained by errors in PIN or in name and they lead to unsuccessful record linkage. Record linkage is important also to define and obtain minimal data sets (for mortality: PIN; family and first name; date and place of birth; gender; residence; date and place of death; underlying and secondary causes of death. The same variables should be considered for hospital discharge diagnosis and admission date and hospital discharge diagnoses);

- exploration about how to obtain the necessary funds to process large administrative files.

Recommendations

As already adopted in many countries, a simplified method, based on a record linkage of hospital discharge diagnoses and death certificates, with validation of a sample of events according to standardised diagnostic criteria is the key recommendation of the EUROCISS Project and might be applied in those countries that do not have a CVD registry. This method uses sources of information and databases currently available in public health services and aims to identify the current numbers of fatal and non-fatal major coronary events [57]. Good technical solutions need to be put in place to maintain patient privacy.

Before stating the simplified method, it is recommended:

- to explore the feasibility of record linkage within hospital records - probabilistic or deterministic approaches based on personal variables or PIN use (within the

same hospital, among hospitals of the area under surveillance, among hospitals at regional level);

- to select a geographical administrative area with a population big enough to provide stable estimates. This means, for example, that a stable population in a representative area of the country with 300 fatal events in the age range 35 to 74 should be chosen in the case of coronary events under the hypothesis of a 2% annual reduction of attack rates in 10 years. If more areas are under surveillance, 300 fatal events should be considered for validation in each separate area [22,31].

E. EUBIROD - European Best Information through Regional Outcomes in Diabetes

The EUropean Best Information through Regional Outcomes in Diabetes-EUBIROD Project is a European diabetes registry, based on integrated health information systems on type 1 and type 2 diabetes from existing national/regional frameworks that uses the BIRO technology; this system automatically generates local statistical reports and safely aggregates data to produce international reports of diabetes indicators.

The Biro project delivered 79 indicators on demographic, clinical and health system characteristics, as well as risk adjusted indicators. 19 countries provided data from different studies based on local data sources. For each indicator, consistency with EUBIROD definition and completeness, was assessed and combined to provide the *Overall Quality Score-OQS*; moreover, for each indicator, some parameters were calculated (e.g. *Percentage Recorded-PR* as percentage of data sources with data item recorded) and used as axes in a graphical plot of Feasibility vs Validity for each data item [49].

The data dictionary (metadata repository) is a central repository of information about data (meaning, relationship to other data, origin, usage, format). The data dictionary is realised through the BIRO common dataset and the EUBIROD survey. The BIRO Common dataset is defined, assessed, and periodically revised by clinical experts, epidemiologists, statisticians, and information technology experts. The EUBIROD survey, conducted across EUBIROD diabetes registers, contributed to assess the consistency of standard definitions with local practices.

Implications

- The BIRO is a system that helps to centralise and aggregate databases in a central server, as an essential element for a secondary use of health data; it is an open source suite of integrated software tools distributed as a complete Linux operating system running on any platform (Windows, Linux);

- BIRO Academy organizes annual residential courses (practical/theory) to learn how to use diabetes data and applications of the BIRO system. Training materials (software, video, demo applications) are available on the academy web site;
- respect of privacy: a novel method of Privacy Impact Assessment is adopted, which ensures complete privacy protection without hampering the information content for public health. In EUBIROD, the clinician who collected the data can correct and control data elements at the basis of diabetes indicators.

Limitations

- In the Common Dataset, event definitions, procedures and methods for data collection are not standardised. Each country adopts its own clinical judgement and sources of information;
- considering that EUBIROD is focused on diabetes, unawareness of diabetes should be included among the indicators of the EUBIROD dataset;
- the parameter assessment criteria involve subjective judgement by the local clinical reviewers and the qualification of reviewers may not be consistent across all centres.

Recommendations

Data dictionaries and data standards can be used to improve quality, consistency and comparability of national information if case definition is standardised (not depending on local practices and clinical reviewers). This procedure may be recommended to assess quality of parameters from different databases.

F. JAMIE - Joint Action on Injury Monitoring in Europe

The JAMIE project was operative from 2011 to 2014 and aimed at having a common hospital-based surveillance system for injury prevention in all MSs by 2015. Such a system should report on external causes of injuries due to accidents and violence as part of the Community Statistics on Public Health.

The project was carried out by a consortium of centres of excellence in injury surveillance based in the EU region (Austria, the Netherlands, Hungary, United Kingdom, Germany). The EuroSafe organisation provided leadership to the project [30].

Implications

Injury data collection efforts should include all acute physical injuries attending Emergency Departments for diagnosis, investigation or treatment, which fall into the nature of injury categories listed in the dataset.

The Injury Database(IDB)-JAMIE data source has been judged as credible and sustainable enough to be included into the ECHI. With respect to injuries, there are a few indicators related to home and leisure injuries - reported by surveys or

from registries - and indicators related to road traffic injuries, work-related injuries, and suicide attempts. The home and leisure injury indicators are defined as injuries that have occurred in and around home, in leisure time and at school, and resulted in an injury that required treatment in a hospital. These data are expected to be provided from national hospital discharge information systems as well as national emergency departments-based injury data, in line with the IDB-JAMIE methodology.
- All countries should implement the core IDB-Full Data Set (FDS) in a representative sample of emergency departments (the IDB FDS is based on a systematic injury surveillance system that collects accident and injury data from selected emergency departments of MS hospitals, existing data sources - such as routine causes of death statistics - hospital discharge registers and data sources specific to injury areas, including road accidents and accidents at work). Where possible, this should be based on injuries from all external causes. In some circumstances, where this is not possible, it may be limited to home and leisure related injuries only.
- Where FDS has not been previously implemented and resources are scarce, each country should implement the FDS in at least one hospital.
- In addition, all countries should widely implement the Minimum Data Set (MDS) (intent, location and setting, activity, and mechanism) unless the FDS in operation provides a sufficiently large and representative sample at a country level. In this case, there is no need for an additional MDS to be collected.

Limitations
- The heavy workload in accident and emergency departments and the limited time for/interest in administrative work put severe pressure on the quality of reporting and completeness of information from the clinical setting.
- The exclusion of many specific and non-specific codes provides a potential for some biases in recording (e.g. road traffic injury is included within the major mechanism category because of the importance of monitoring and supporting road traffic injuries in almost all settings. Of course, road traffic injuries occur due to a variety of mechanisms including cutting/piercing, burns and sheering stresses, but the vast majority are due to blunt force from contact with hard objects).

G. EUROHOPE European Health Care Outcomes, Performance and Efficiency

EuroHOPE-European Health Care Outcomes, Performance and Efficiency- evaluates the performance of European healthcare systems in terms of outcomes, quality, use of resources and costs; five key public health problems/diseases are evaluated: acute myocardial infarction, stroke, hip fracture, breast cancer, and low-birthweight infants. Moreover, by using data from Nordic hospitals, it is also possible to measure the productivity related to the quality of care [44]. In addition to developing performance indicators, the project makes studies aiming to explain

and understand the reasons for national, regional and hospital level differences in health care performance [58].
Currently national, regional and hospital level indicators have been calculated from Finland, Denmark, Hungary, Italy, Norway and Sweden. The episode-based approach have been extended to include primary health care and social services in a pilot study using data from Copenhagen, Helsinki Oslo and Stockholm.[58]

EuroHOPE is based on analysing the progress of a disease, with specific interest in the role of health services and health care policy as determinants of the progress. The main idea of the approach is to analyse the performance by using detailed data pertaining to specific health conditions and shed light on the interconnected aspects (i.e. financing, organisational structures, medical technology choices) that are responsible for health system performance, health outcomes and expenditure. By making use of the available databases, the project updates and further develops research infrastructure, with the aim of evaluating the performance of health care systems in terms of outcomes, quality, use of resources, and cost. This includes maintaining and updating work protocols for selected conditions. The protocols include inclusion and exclusion criteria, definition of the cycle of care and co-morbidities used in risk adjustment, and describe the specification of process, utilization, cost and outcome measures.

A performance measure must be carefully constructed and be appropriate for multinational comparisons. Performance indicators will vary according to the type of hospital, regional or individual level variations or random variations. The focus of our interest is variation at hospital level, regional level and country level. In EuroHOPE, the effect on the measures caused by cross-country heterogeneity in patient casemix is reduced by using risk-adjustment methods for individual-level data [44,45,59].

Limitations

-Usefulness of the approach depends on comprehensive register data being available and the possibility of linking the hospital discharge register to other databases.

-For some of participating countries, the formal and technical processes of having access to data are too slow and cumbersome.

-The differences in coding practices across countries and the quality of data is not always comparable.

Recommendations

-The methodological framework developed in EuroHope provides a solid starting point for further elaborating an international health care performance assessment toolkit.

-It is widely recognized that the ability to follow patients as they progress through the health care system to death, including primary health care, hospital and other specialized treatments, long-term care, home care and hospice care, is essential to health care quality and performance assessment. The pilot study made in capital areas of Nordic countries extending episode based approach to include primary health care and social services indicates that the methodology developed in the project can be extended to new countries and conditions. For example, an electronic patient record system (including all health care activities) is under development in many countries and will give new, path-breaking possibilities for the development of the disease-based approach. This requires data using standardised and internationally comparable definitions of activities and classifications describing the treatments (i.e. diagnosis, procedures) are nationally available for research and thus enable evaluation of the performance across countries, regions and producers.

H. COPHES Consortium to Perform Human Biomonitoring on a European Scale

The protocols developed and harmonised by COPHES allow the collection of comparable human biomonitoring data throughout Europe; DEMOCOPHES is the feasibility study, launched one year after COPHES to collect information and tests on major determinants of exposure in Europe, and to establish protocols for the translation of results into concrete recommendations.

The study measured biomarkers for mercury, cadmium, phthalates, bisphenol, environmental tobacco smoke in human hair and urine from 120 mother-child pairs in 17 countries (4000 samples) [33-38].

Implications

COPHES and DEMOCOPHES have been able to demonstrate that a more coordinated and harmonized approach to human biomonitoring is possible in Europe, that it can become an important tool to monitor the exposure of Europeans to chemical substances and address potential health effects that may derive from this exposure. COPHES/DEMOCOPHES results demonstrated that harmonization is possible, but further capacity building is needed to establish the networks and infrastructure in EU countries; a European database on the distribution of the chemical burden in the population is available; it will allow to follow the trends in the population and evaluate the effectiveness of regulatory measures, such as a

ban on certain chemicals, or smoking regulations. At the same time, an improved comparability of European data will allow to evaluate the gradients in human exposure throughout Europe and facilitate the elaboration of guidance values and the identification of potential high exposure of subgroup of populations, thus helping to produce targeted measures.

Limitations

Difficulties exist in biomonitoring approaches to adequately link the relevant health outcomes, as these may take place decades later.

Recommendations

Longitudinal studies are essential to place the information needs in the very costly area of longitudinal research; while monitoring exposure would be relatively simple, given good knowledge on the effects to be expected.

I. ECHO-Health – European Collaboration for Healthcare Optimization

The goal of ECHO was to describe and analyse health care system performance in terms of healthcare utilization, equity of access to effective care, and quality and efficiency. ECHO identified unwarranted variation in healthcare service delivery within and across countries at different levels of analysis: hospital, healthcare area and region [60,61].
The ECHO project was an international effort to gather healthcare information from several European countries within a single data warehouse (DWH). The information collected specifically concerned patient-level data from hospital admissions and demographic and socio-economic information at the geographic level. The project also aimed at supplying information at both hospital and geographic levels. The countries participating in the project were Austria, Denmark, England, Portugal, Slovenia and Spain. ECHO was built upon routinely collected administrative data.
ECHO used two complementary methodological approaches, a geographical and a hospital-specific one, to evaluate more than 40 validated indicators, classified as calibrators, cardiovascular care, orthopaedics, patient safety, low value care and potentially avoidable hospitalizations.

Implications
- Several minimum information requirements (i.e., a set of variables comprising the ECHO core dataset and used to integrate the original hospital administrative datasets into a single and coherent relational database), were imposed upon original raw data sources before inserting them into the ECHO data model. Some of these variables were patient attributes (e.g., age, sex, diagnoses and procedures),

episode attributes (e.g., type of discharge or hospital of treatment), and attributes allowing patient geo-allocation. The ECHO data model accounts for approximately 98% of the hospitalization episodes occurred in the participant countries from 2002 to 2009, (with the exception of Slovenia, who measured hospitalisations over the 2005-2009 period), adding up to 192 million episodes. [62]
- Data quality assessment was based on the Quality Assessment Framework (QAF) [2]. Also, the main data quality dimensions assessed were: coherence, coverage, relevance, internal reliability/comparability, and accuracy.
- Indicators were chosen based on their validation at international level in prior studies, including some conducted by the Coordinators' group for Spain. The data model was specifically designed for this project, and quality dimensions of the relational data model were also tested.
- External validity was assessed by a mixed approach that evaluated face-validation with partners from the different countries and compared projects' findings with results from national statistics and other published studies.
- Face-validation was also tested at large policy-dialogues with healthcare managers and decision-makers from each country.

Limitations
- Hospitalization episodes do not allow to consider death after discharge. Additionally, re-admissions could not be analysed since patient identification was not provided by all the participant countries.
- The information available on contextual factors, namely hospital supply and population socioeconomic status, was heterogeneous amongst participant countries.
- Geographical analysis requires geo-localization of the assisted population based on information about the healthcare area, ZIP code or other geographical info that was usually incomplete.
- To ensure comparability across countries (at any level of disaggregation), several requirements were imposed, including a common definition of what had to be considered a valid hospitalization episode and how to build a new episode from the information available for each country.
- Some countries (Denmark, England and Portugal) didn't have administrative healthcare areas that allowed comparison with other countries, thus it was necessary to build administrative areas based on a previous analysis of the healthcare administration levels and population distribution [63].
- ECHO did not address some of the dimensions of usual performance assessment frameworks, such as appropriateness, system responsiveness, timeliness, or patients' satisfaction".

Recommendations:
The ECHO information system quality report compiled a list of recommendations for an optimal use of the information stored in the DWH, highlighting specific

issues that should be taken into account to ensure proper interpretation of ECHO performance indicators.
More information about the quality methods developed by ECHO can be found in the Handbook on methodology: ECHO information system quality report [62].

IX. CONCLUSION AND RECOMMENDATIONS

Without good data, quality of indicators, quality of studies and therefore decisions on planning and evaluation of preventive programmes, health care delivery, resource allocation, and research are severely impaired.
Quality should concern data, data sources and indicators. This report contains an overview of the quality methods related to these three topics.

How to improve data quality

The first step to plan and organize data collection in surveys and registries is to **prepare the manual of operations**. It includes all specific information on methods and procedures, such as the definition of target population, sampling, measurements, questionnaire, communication to participants, data analysis and data storing. If the aim of data collection is also to build one or more indicators, it is important to describe **how data must be processed and computed to build indicators**. All steps included in the manual of operations should be checked, and quality of performance and collected data should be tested. In this context, country-specific conditions and health monitoring needs should be considered.

Studies based on administrative data, in which two or more datasets are linked, require standardisation of the event definition, harmonization of data for a reliable and comparable definition, description of the sources of information and procedures to aggregate and elaborate data in order to generate indicators.

The following steps are crucial to improve data quality:

A. TRAINING AND TESTING

The staff that performs measurements or collects data should be qualified through training and testing. After a complete information on the study (objectives, procedures and methods for measurements), usually given by an expert, the staff should attend practical sessions. In this activity, the staff firstly observes an expert who performs standardised procedures and methods performed for measurement and/or questionnaire administration, or for results codification and computer input; then the staff replicates the same procedures under the supervision of the expert. The staff is trained up to when it can perform the procedures as indicated in the manual of operations. The testing process is the agreement of

staff performance results and the established values. Training and testing can improve data reliability.
E.g. *Method to train and test the staff for measuring blood pressure before the health examination survey and to assess the quality of data collected*. This training includes an explanation of the reason why standardised blood pressure measurements are needed, the conditions requested to the examined person for the measurement (abstain from doing heavy physical activity, eating, smoking and avoid exposure to cold temperature for one hour before measurement), number of measurements needed (two/three consecutive measurements), position of the personnel and position of the examined person (sitting position), key steps in the measurement procedure (selection of the cuff, high level to inflate, deflate the cuff at a rate of 2mmHg per second), how results should be recorded (as a mean of two/three measurements); as well as how the results should be explained to the participant.
Practical training includes an adequate number of measurements under the supervision of an expert and feedback sessions to discuss errors during the measurements. The double stethoscope is used to check the readings of blood pressure measurement; difference of more than two mmHg between trainer and trainee is not allowed. At regular intervals, the following quality checks are needed: distribution of last digits for systolic and diastolic measurements to test the accuracy of measurements, and proportion of identical measurements in the same participant in order to test that three measurements of blood pressure are really done. Site visits during the field work and audits are important to maintain quality control.

B. DATA COLLECTION

Procedures, methods and tools used during the study have an impact on the final results of the study. They should be standardised, comply with the purpose of the study, match the population under evaluation, and be completed. They should provide good quality data, do not overload the participant (in the case of an ad hoc survey), comply with ethical and data-protection requirements, and have limited costs. Rules for, and comments on the implementation of the data collection should be fixed in a written form and made available to the data collection personnel [55]. An *ad hoc* survey should be based on measurements and/or questionnaires, while population based registries should be based on integration or linkage of several data sources (e.g. hospital discharges, death certificates, drug prescriptions, etc.).

Examples of quality control of blood pressure measurements during the European health examination survey are shown in table 4 [55]:
"In nine out of 12 surveys, more than one cuff size was available. Arm circumference was measured in ten surveys. When comparing measured arm circumferences to the size of the cuff used for the measurement, the miss-cuffing (use of too small or too large cuff)

was observed only in 1-5% of the subjects, except in one survey where only one cuff was available and 20% of subjects were miss-cuffed. In three surveys which did not measure arm circumference, the occurrence of miss-cuffing would have been more likely to happen, especially if the used cuffs did not have markings to indicate correctness of the cuffs for the specific arm circumference. In all these three surveys, cuffs with indicators to assess the correctness of the cuff size were used. The proportion of identical sequential measurements was lower between the first and the second measurement than between the second and the third measurement for both systolic and diastolic blood pressure in three surveys using simple mercury sphygmomanometers. In each survey, the proportion of identical readings was higher for diastolic than for systolic blood pressure. Overall, the proportion of identical sequential measurements was high only in two surveys (28% or over) using simple mercury sphygmomanometer."

Table 4: EHES: Blood pressure measurements in 12 surveys. Recording of cuff size used and measured arm circumference, proportion of miss-cuffed subject, and proportion of identical readings between 1st-2nd and 2nd-3rd measurements.

Pilot survey	Recorded cuff size used	Arm circumference measured and recorded (M = measured, not recorded, B = measured and recorded, N = Not measured)	Mean arm circumference (min,max)	Proportion of miss-cuffed subject (based on optimal arm circumference reported on the cuff)	Proportion of identical readings between 1st and 2nd measurement		Proportion of identical readings between 2nd and 3rd measurement	
					SBP	DBP	SBP	DBP
A	No	B	30.8 cm (24.5,43.5)	#	8%	29%	16%	28%
B	Yes	B	31.6 cm (23.0,48.0)	1%	10%	29%	31%	41%
C	Yes	B	28.3 cm (20.0, 40.0)	0%	10%	13%	19%	19%
D	Yes	N	§	#	4%	10%	7%	12%
E	Yes	B	29.9 cm (22.0,41.5)	0%	5%	9%	6%	11%
F	Yes	B	28.9 cm (21.0,41.5)	2%	3%	9%	9%	11%
G	Yes	M	§	#	7%	12%	8%	13%
H	Yes	B	30.1 cm (21.0,48.0)	20%	3%	8%	5%	9%
I	Yes	B	30.9 cm (22.0,46.0)	0%	8%	10%	8%	10%
J	Yes	N	§	#	6%	14%	9%	15%
K	Yes	B	30.4 cm (21.5,43.3)	5%	4%	10%	7%	10%
L	Yes	B	30.4 cm (24.0,40.0)	1%	5%	9%	5%	11%

§ Arm circumference not measured, # Not possible to calculate.

Source: Tolonen H, Koponen P, Naska A, Männistö S, Broda G, , Kuulasmaa K7; EHES Pilot Project. Challenges in standardisation of blood pressure measurement at the population level. BMC Med Res Methodol. 2015 Apr 10;15:33 [55].

C. DATA INPUT AND DATA DELIVERY

Data collected by questionnaires, measurements or routinely sources sometimes need to be codified (this process assigns a value to a response) and entered into a computer system. Usually the data entries of questionnaires and routinely collected sources are computer assisted and are automatically recorded and codified in an electronic way. This method reduces cost and enhances accuracy of the results in comparison to a manual codification and insertion of data.

Coding errors could be minimized by training the personnel and requiring that the same data are inserted by more than one person. Double entry of subsamples can be considered.

To identify errors due to wrong coding or wrong data entry, preliminary quality checks should be performed and should assess:

- *for quantitative variables:* format, plausibility of values, anomalous data, range, distribution, number of missing values, and consistency with other related variables;
- *for qualitative variables*: format, allowed codes, distribution, missing values, and consistency with other related variables.

The identification and correction of the errors due to wrong coding or wrong data entry could be done by verifying the original data stored on paper or electronic support, or by deciding to report the wrong values as missing values.

Arbitrary decisions and interventions on databases should be as rare as possible, and completeness of data should be reported in the results of the statistical analysis.

In studies where data from different sources are included, data may need to be transferred to a central entity. In this case, a clear definition of the data delivery procedures is desirable. For example, in the Data Delivery Guidelines for the European Health Interview Survey, the data transmission way is described in detail, including the validation rules to be applied to the delivered data. [64]

The validation rules define how to perform checks for allowed codes and values, skip checks, and control consistency .

D. DEFINITION OF NEW VARIABLES OR INDICATORS

The definition of new variables or indicators should be clear and precise. Changes to variable values or the building of new variables are to be documented in writing in each individual case [46]. The selection process of a new variable's definition depends on the nature of the data collected, the nature of the new variable and the purpose to be served by the new variable. The latter could be a simple transformation of a continuous variable by using a different unit of measurement, or a new codification of a quality variable by aggregating possible answers, or an aggregation of quantitative and qualitative variables. Great attention should be paid to the possible values of the new variables, and particular care is needed in the use of missing values, especially when more variables are aggregated.

An example of aggregation of quantitative and qualitative variables is the definition of hypertension (yes/no), in which three variables collected during the

survey may be considered: systolic and diastolic blood pressure (quantitative variables) and use of antihypertensive treatment (qualitative variables). The presence of hypertension may be defined as systolic blood pressure≥140 mmHg or diastolic blood pressure≥ 90mmHg, or under specific treatment. Hypertension could be defined in several ways; the following three ways are provided as examples:

- systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg or under specific treatment, excluding those subjects with one or more missing variables;
- systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg or under specific treatment, excluding those subjects with all three missing variables;
- Systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg not considering specific treatment.

The definition of hypertension and use of missing values is crucial for the estimation of the prevalence of hypertensive people.

When data of two or more surveys/databases are pooled, it is crucial to know which procedures and methods have been used to measure blood pressure, which questions have been used to record treatments and which the definition of hypertension has been used (including use of missing values). When two or more surveys/databases are aggregated, it is better to include all variables collected during the survey (in the mentioned example, blood pressure measurements and use of antihypertensive medications) rather than to report only the new variable "hypertension", in order to be sure to use a common definition for all involved databases.

In a population-based register, an example of aggregation of variables collected within the study could be the event validation process. This process is based on a precise and complex flow chart, that starts with a record linkage of hospital discharges and death certificates, and allows to build the "current event" (yes/no) variable, and then to start a standardised validation process of this information to build the new "validated event" (yes/no) variable. This example underlines the importance of collecting and checking completeness and coherence of data collected within the registry and of specifying a precise process to build new variables (current event and validated event).

Recommendations

The assessment of data quality is crucial to assess the reliability and comparability of data among countries and across regions, to monitor time trends and build health information systems at national and European level.

The first step to provide quality data collection in surveys and registries is to **prepare the manual of operations,** which includes all specific information on methods and procedures, as, for instance, definition of target population, sampling, measurements, questionnaire, communication to participants, data analysis and data storing. If data collection also aims at building one or more indicators, it is important to describe **how data must be processed and computed to build indicators**.

After the survey is performed and registry and administrative data statistics are collected, the subsequent step involves the preparation of a **data quality report** with detailed description of data, data sources, their size and characteristics, the process used to compute indicators, and all related quality measures and checks, which are fundamental for aggregation, harmonization and comparison of indicators. Quality measurement should be part of any statistical process and should not be a separate activity carried out after the statistics are produced or when users need it. This is not only done for the sake of cost and time efficiency, but also for purposes of quality improvement.

The third step is a prompt **feedback to the personnel** involved in collecting, harmonizing, and processing data, as this is the best way to improve data quality.

Expert Group

This report is based on the work of the many experts who contributed to its development.

The following experts performed the literature review, collected the information and participated as writing group:

Simona Giampaoli (simona.giampaoli@iss.it), Istituto Superiore di Sanità, Italy

Peter Achterbeg (peter.achterberg@rivm.nl), National Institute for Public Health and the Environment , Netherlands

Enrique Bernal-Delgado (ebernal.iacs@aregon.es), Institue of Health Research, Spain

Petronille Bogaert (petronille.bogaert@wiv-isp.be), Scientific Institute of Public Health, Belgium

Flavia Carle (f.carle @univpm.it), Università Politecnica delle Marche, Italy

Anna Di Lonardo (anna.dilonardo@iss.it), Istituto Superiore di Sanità, Italy

Kris-Doggen (kris.doggen@wiv-isp.be), Belgian Scientific Institute for Public Health, Belgium

Chiara Donfrancesco (chiara.donfrancesco@iss.it), Istituto Superiore di Sanità, Italy

Angela Fehr (FehrA@rki.de), Robert Koch Institute, Germany

Lidia Gargiulo (gargiulo@istat.it), Italian Institute of Statistics, Italy

Häkkinen Unto (unto.hakkinen@thl.fi), National Institute for Health and Welfare, Finland

Sabrina Hense (HenseS@rki.de), Robert Koch Institute, Germany

Maria M. Hofmarcher-Holzhacker (maria.hofmarcher@healthsystemintelligence.eu,) HealthSystemIntelligence, Austria

Laura Iannucci (iannucci@istat.it), Italian Institute of Statistics, Italy

Luigi Palmieri (luigi.palmieri@iss.it), Istituto Superiore di Sanità, Italy

Hanna Tolonen (hanna.tolonen@thl.fi), National Institute for Health and Welfare, Finland

Serena Vannucchi (serena.vannucchi@iss.it), Istituto Superiore di Sanità, Italy

Jennifer Zeitlin (jennifer.zeitlin@inserm.fr), Institut national de la santé et de la recherche médicale (INSERM), France

X. REFERENCES

1. https://stats.oecd.org/glossary/detail.asp?ID=2620
2. Bergdahl M, Ehling M, Elvers E, Földesi E, Körner T, Kron A, Lohauß P, Mag K, Morais V, Nimmergut A, Sæbø HV, Timm U, Zilhão MJ. Handbook on Data QualityAssessment Methods and Tools. EUROSTAT. 2007; M Ehling and T Körner (eds) pg 1-139
3. ESS handbook for quality reports. EUROSTAT Manuals and guidelines. European Statistical System. 2014 pg 1-164
4. Signore M, Brancato G, Carbini R, D'Orazio M, Simeoni G. Linee guida per la qualità dei processi statistici. 2012; ISTAT.pg 1-52
5. AHRQ Quality Indicators—Guide to Prevention Quality Indicators: Hospital admission for Ambulatory Care Sensitive Conditions. Rockville, MD: Agency for Healthcare Research and Quality. Revision 3. (January 9, 2004). AHRQ Pub. No. 02-R0203
6. EUROCISS- European Cardiovascular Indicators Surveillance Set, FINAL REPORT; 14 February 2008; http://www.cuore.iss.it/eurociss/en/projectII_report2007.asp
7. EHES-http://www.ehes.info/
8. Non Communicable Diseases Risk Factor Collaboration. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19 million participants. Lancet 2016; 387: 1377–96
9. Non Communicable Diseases Risk Factor Collaboration 2. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants. Lancet 2016; 387: 1513–30
10. Eurostat/A4/Quality/03/General/Definition Working Group "Assessment of quality in statistics" Sixth meeting Luxembourg, 2-3 October 2003 ITEM 4.2 Methodological documents - definition of quality in statistics
11. http://ec.europa.eu/eurostat/ramon/nomenclatures/index
12. OECD :glossary of statistics terms: htpps://stats.oecd.org
13. http://www.nationaalkompas.nl/object_binary/o14707_ECHIM-final-report_Part-II_PDF.pdf
14. Tolonen H, Koponen P, Mindell JS et al. Under-estimation of obesity, hypertension and high cholesterol by sef-reported data: Comparison of self-reported information and objective meaures from health examination surveys. Eur J Public Health 2014; 24:940-947
15. www.cuore.iss.it
16. http://www.epicentro.iss.it/passi/dati/sovrappeso.asp
17. http://www.istat.it/it/files/2016/07/Fattori-di-rischio_salute_def.pdf?title=Fattori+di+rischio+per+la+salute++-+26%2Flug%2F2016+-+Testo+integrale.pdf
18. Donfrancesco C, Lo Noce C, Brignoli O, Riccardi G, Ciccarelli P, Dima F, Palmieri L, Giampaoli S, Oprg OP. Italian network for obesity and cardiovascular disease surveillance: a pilot project. BMC Family Practice 2008; 9(1): 53
19. C. Batini, M. Scannapieca. Data Quality: Concepts, Methodologies and Techniques, Springer, Heidelberg 2006

20. Tunstall-Pedoe H, Kuulasmaa K, Tolonen H, Davidson M, Mendis S with 64 other contributors for The WHO MONICA Project. MONICA Monograph and Multimedia Sourcebook. Geneva: World Health Organization; 2003
21. Porta M. (2008) A dictionary of Epidemiology, 5th ed., Oxford University Press, New York
22. Madsen M et al: Population based registry of acute myocardial infarction : manual of operations. Cardiovascular surveillance: The EUROCISS Project. Europ J Cardiov Prev Rehabilit 2007; 14:3-22
23. http://ec.europa.eu/eurostat/statisticsexplained/index.php/Glossary
24. http://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey
25. Tolonen H (ed.) (2013a) EHES Manual, Part A. Planning and preparation of the survey. National Institute for Health and Welfare. Directions 2013_001. URN:ISBN:978-952-245-842-1, URL:http://url.fi/URN:ISBN:978-952-245-842-1
26. Tolonen H (ed.) (2013b) EHES Manual, Part B. Fieldwork procedures. National Institute for Health and Welfare. Directions 2013_002. URN:ISBN:978-952-245-843-8,URL:http://urn.fi/URN:ISBN:978-952-245-843-8
27. Commission Regulation (EU) No 141/2013 of 19 February 2013 implementing Regulation (EC) No 1338/2008 of the European Parliament and of the Council on Community statistics on public health and health and safety at work, as regards statistics based on the European Health Interview Survey (EHIS). Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013R0141&from=EN
28. Eurostat (2013). European Health Interview Survey (EHIS wave 2). Methodological manual. European Union. Available at: http://ec.europa.eu/eurostat/documents/3859598/5926729/KS-RA-13-018-EN.PDF/26c7ea80-01d8-420e-bdc6-e9d5f6578e7c
29. Nwaru BI, Mukherjee M, Gupta RP, Farr A, Heaven M, Stoddart A, Bandyopadhyay A, Fitzsimmons D, Shields M, Phillips C, Chamberlain G, Fischbacher C, Dibben C, Aftab C, Simpson CR, Lyons R, Strachan D, Davies G, McKinstry B, Sheikh A. Challenges of harmonising data from UK national health surveys: a case study of attempts to estimate the UK prevalence of asthma. J R Soc Med 2015. OnlineFirst, published on October 2, 2015 as doi:10.1177/0141076815600909
30. EuroSafe (2014). IDB-JAMIE Full data set (IDB-FDS) Data Dictionary. Available at: http://www.eurosafe.eu.com/uploads/inline-files/IDB%20JAMIE%20FDS%20Data%20Dictionary%20MAR14.pdf
31. Giampaoli S, Hammar N, Adany R et al. Population-based register of stroke: manual of operations. Eur J Card Prev and Rehabil 2007; 14 (Suppl 3): S23-S41
32. Primatesta P, Allender S, Ciccarelli P et al. (2007) Cardiovascular surveys: manual of operations. Eur J Card Prev and Rehabil 2007; 14 (Suppl 3): S43-S61
33. Becker K, Seiwert M, Casteleyn L, et al. (2014) A systematic approach for designing a HBM pilot study for Europe. Int J Hyg Environ Health, 217(2-3): 312-322. doi:10.1016/j.ijheh.2013.07.00
34. Casteleyn L, Dumez B, Becker K et al. (2015) A pilot study on the feasibility of European harmonized human biomonitoring: Strategies towards a

common approach, challenges and opportunities. Environ Res, 141: 3-14. doi:10.1016/j.envres.2014.10.028
35. Esteban M, Schindler BK, Jimenez JA et al. (2015) Mercury analysis in hair: Comparability and quality assessment within the transnational COPHES/DEMOCOPHES project. Environ Res, 141: 24-30. doi:10.1016/j.envres.2014.11.014
36. Exley K, Cano N, Aerts D et al. (2015) Communication in a Human biomonitoring study: Focus group work, public engagement and lessons learned in 17 European countries. Environ Res, 141: 31-41. doi:10.1016/j.envres.2014.12.00
37. Fiddicke U, Becker K, Schwedler G et al. (2015). Lessons learnt on recruitment and fieldwork from a pilot European human biomonitoring survey. Environ Res, 141: 15-23. doi:10.1016/j.envres.2014.08.039
38. Schindler BK, Esteban M, Koch HM et al. (2014) The European COPHES/DEMOCOPHES project: towards transnational comparability and reliability of human biomonitoring results. Int J Hyg Environ Health, 2017(6): 653-661. doi:10.1016/j.ijheh.2013.12.002.
39. Gissler M, Mohangoo AD, Blondel B, Chalmers J, Macfarlane A, Gaizauskiene A, Gatt M, Lack N, Sakkeus L, Zeitlin J; Euro-Peristat Group. Perinatal health monitoring in Europe: results from the EURO-PERISTAT project. Inform Health Soc Care. 2010 Mar;35(2):64-79. doi: 10.3109/17538157.2010.492923
40. Euro-Peristat (2012) Euro-Peristat list of indicator, updated 2012. Available at: http://www.europeristat.com/images/doc/updated%20indicator%20list.pdf
41. www.qualityindicators.ahrq.gov
42. www.atlasvpm.org
43. http://www.echo-health.eu/handbook/infrastructure.html
44. Häkkinen U, Iversen T, Peltola M, Seppälä T, Malmivaara A, Belicza É, Heijink R, Fattore G, Numerato D, Medin E, Rehnberg C. Health care performance comparison using a disease-based approach: the EuroHOPE project. Health Policy 2013; 112(1-2): 100-9
45. http://www.eurohope.info/
46. http://ec.europa.eu/health/indicators/echi/index_en.htm
47. http://www.healthdatanavigator.eu
48. Carinci F, Di Iorio CT, Massi Benedetti M, Standardized information exchange in diabetes: integrated registries for governance, research and clinical practice, in Bruttomesso D, Grassi G (eds), Technological Advancements in the Treatment of Type 1 Diabetes, 2015; (Frontiers in Diabetes, vol 24), p.236-249, Basel, Karger
49. Cunningham SG, Carinci F, Brillante M, Leese GP, McAlpine RR, Azzopardi J, Beck P, Bratina N, Boucquet V, Doggen K, Jarosz-Chobot PK, Jecht M, Lindblad U, Moulton T, Metelko Ž, Nagy A, Olympios G, Pruna S, Skeie S, Storms F, Di Iorio CT and Massi Benedetti M, Defining a European diabetes data dictionary for clinical audit and healthcare delivery: core standards of the EUBIROD project, Methods of Information in Medicine, 2015 Dec 15;55
50. Mohangoo AD, Buitendijk SE, Szamotulska K, Chalmers J, Irgens LM, Bolumar F, Nijhuis JG, Zeitlin J. Gestational age patterns of fetal and neonatal

mortality in europe: results from the Euro-Peristat project. PloS one. 2011; 6(11):e24727
51. Delnord M, Hindori-Mohangoo AD, Smith LK, Szamotulska K, Richards JL, Deb-Rinker P, Rouleau J, Velebil P, Zile I, Sakkeus L, Gissler M, Morisaki N, Dolan SM, Kramer MR, Kramer MS, Zeitlin J, Euro-Peristat Scientific C. Variations in very preterm births rates in 30 high-income countries: are valid international comparisons possible using routine data? BJOG. 2016
52. http://www.euroreach.net/compendium
53. The Health Data Navigator toolkit www.healthdatanavigator.eu
54. Tolonen H, Mki-Opas JS, Trichopoulou A, Naska A, Männistö S, Giampaoli S, Kuulasmaa K, Koponen P for the EHES Pilot Project. Standardization of physical measurements in European health examination surveys-experience from the site visits. Eur J Public Health 2017 doi: 10.1093/eurpub/ckw271
55. Tolonen H Koponen P, Naska A, Männistö S, Broda G, , Kuulasmaa K7; EHES Pilot Project. Challenges in standardization of blood pressure measurement at the population level. BMC Med Res Methodol. 2015; 15(1):33 doi:10.1186/s12874-015-0020-3
56. Tolonen H, Hoponen P, Mindell J, Männistö S, Kuulasmaa K. European Health Examination survey- towards a sustainable monitoring system Eur J Public Health. 2014; 24(2):338-44. doi: 10.1093/eurpub/ckt107.
57. Ferrario M, Cesana G, Vanuzzo D, Pilotto L, Sega R, Chiodini P, Giampaoli S. Surveillance of ischaemic heart disease: results from the Italian MONICA populations. International Journal of Epidemiology. 2001; 30: S23-S29
58. Häkkinen U, Iversen T, Peltola M, Rehnberg C, Seppälä T. 2015. Towards explaining international differences in health care performance: Results of the EuroHOPE Project. Health Economics 24(Suppl. 2): 1–4.
59. Moger TA, and Peltola M. 2014. Risk adjustment of health-care performance measures in a multinational register-based study: A pragmatic approach to a complicated topic. Sage Open Medicine, 2, 1-11.
60. http://echo-health.eu/ (Accessed October 2016)
61. Bernal-Delgado E, Christiansen T, Bloor K, et al. ECHO: health care performance assessment in several European health systems. Eur J Public Health. 2015 Feb;25 Suppl 1:3-7. doi: 10.1093/eurpub/cku219.
62. European Collaboration for Healthcare Optimization (ECHO) www.echo-health.eu .Zaragoza (Spain): Instituto Aragonés de Ciencias de la Salud - Instituto Investigación Sanitaria Aragón; c2011. Estupiñán F, Baixauli C, Bernal-Delgado E on behalf of the ECHO consortium. Handbook on methodology: ECHO information system quality report; 2014 Apr 27; available from: http://www.echo-health.eu/handbook/ infrastructure.htm (Accessed October 2016)
63. Thygesen LC, Baixauli-Pérez C, Librero-López J, et al. Comparing variation across European countries: building geographical areas to provide sounder estimates. Eur J Public Health. 2015; 25 Suppl 1:8-14. doi: 10.1093/eurpub/cku229
64. https://circabc.europa.eu/webdav/CircaBC/ESTAT/healthtf/Library/technicalsgroupshis/2013_%2026%20February/Presentations/TG%20HIS%202013%20-%204_5.pdf.