# D8.1-Health Information System development: data collection and quality assurance for a common health information system

## Work package 8. Tools and methods for health information support

### Task 8.1 Generating knowledge on data collection methods, and availability and accessibility of health information

National Institute of Health (ISS), Italy

Authored by Brigid Unim, Luigi Palmieri
in collaboration with WP 8 partners
WP 8 Lead: National Institute of Health (ISS), Italy
Joint Action Coordination:
Sciensano | Rue Juliette Wytsmanstraat 14 |
1050 Brussels | Belgium| e-mail : infact.coordination@sciensano.be|
Website: www.inf-act.eu|Twitter: @JA_InfAct

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

# Table of Contents

## Executive summary

The Report on health data collection methods and procedures is the first Deliverable of the Work Package 8 (WP8) of the Joint Action on Health Information (InfAct). The report outlines the results of WP conducted under Task 8.1 (T8.1) - Generating knowledge on data collection methods, and availability and accessibility of health information. The aim of T8.1 is to identify data collection methods, quality assurance, and availability and accessibility of health information across Member States (MSs). To that regard, a review of international organizations and selected EU research networks was implemented and a structured questionnaire was developed and administered to all MSs' representatives.

Health-related data are collected from a variety of sources such as population-based registries, health interview and examination surveys, longitudinal studies, administrative healthcare records, e-health solutions, and more. Data is collected for different purposes, including population health monitoring/public health surveillance and health system performance assessment. Health monitoring data provide the main information for the description of population health status, while performance measurement seeks to monitor, evaluate and communicate the extent to which various aspects of the health system meet the key objectives.

The findings of the cross-sectional study underline the gap in health data and information availability, accessibility, comparability or reusability for research purposes and policy making. In fact, only 30% of the identified projects share data with other EU projects or research networks, limiting the use of health data in and across EU countries. The findings of the study will facilitate the assessment of health inequalities across EU countries in terms of quality, availability, accessibility and comparability of health data and information. It will also facilitate sharing and dissemination of standardised and comparable health data collections, by providing research results to the InfAct web-based platform, a one-stop-shop for health information research in EU.

# Health data collection methods and procedures

## I. Introduction

The main outcome expected from InfAct, a joint action (JA) on health information (HI), is a sustainable solid infrastructure on EU HI to improve the availability of comparable, robust and policy relevant data on health status and health system performance. Through country collaboration, the JA aims to streamline HI activities, reduce the data collection burden and works for a sustainable and robust data collection in Europe that facilitates and supports country knowledge, health research and policy making.

Nationally, health-related data are collected from a variety of sources such as population-based registries, health interview and examination surveys, longitudinal studies, administrative healthcare records, e-health solutions, and more. Data is collected for different purposes, including population health monitoring/public health surveillance and health system performance assessment (HSPA). Most of these data are not included in international databases such as the World Health Organization (WHO), Organization for Economic Co-operation and Development (OECD) or the European Statistical Office (Eurostat), limiting their use for research, policy, international benchmarking and comparisons.

Health monitoring data provide the main information for the description of population health status. Monitoring is an intermittent or episodic performance and analysis of measurements aimed at detecting changes in the health status of populations or in the physical or social events [1]. On the contrary, surveillance is a continuous process that requires three functions in this sequence: i) data collection; ii) analysis and interpretations; and iii) decision making. The final phase in the surveillance chain is the application of information to health promotion and to disease prevention and control. Public health surveillance is defined here, as the ongoing systematic collection, analysis, and interpretation of health data, essential to the planning, implementation, and evaluation of public health practice, closely integrated to the dissemination of these data to those who need to know and linked to prevention and control [2].

Performance measurement seeks to monitor, evaluate and communicate the extent to which various aspects of the health system meet the key objectives. There is consensus among members of the Committee on the National Quality Report on Health Care Delivery [3] and clinical experts participating in the OECD Health Care Quality Indicator Project [4]

that those objectives can be summarized as: i) health conferred on citizens by the health system; ii) responsiveness to individual needs and preferences of patients; iii) financial protection offered by the health system; and iv) productivity of utilization of health resources. A healthcare system should also fulfil other criteria such as equity on access, effectiveness, quality and safety, and allocative efficiency [5].

Standardization for data collection for monitoring/surveillance or HSPA is required to ensure comparability of the results. Comparability is often restricted by differences in definitions, used collection methods and tools, and varying uses of classifications. Standardization procedures ensure that three criteria are met: i) the aims of data collection are made explicit and all necessary and pertinent information are collected; ii) data are collected using the same method; iii) the same definitions are used. Standardization is also time efficient and essential for comparing population groups, geographic areas, or trends over long periods of time [6]. Some examples of standardized data collection are: i) Eurostat health statistics collected from different sources under specific regulations [7]; and ii) the countries participating in the European Health Examination Surveys (EHES) research network also follow standardized data collection methods and procedures [8]. Standardization of metadata is also important in health information systems describing health data. For example, the main reference metadata-reporting standards used by Eurostat [9] are: i) SIMS (Single Integrated Metadata Structure); ii) ESMS (Euro SDMX Metadata Structure); iii) ESMS-IP (Euro SDMX Metadata Structure – Indicator Profile); and iv) ESQRS (ESS Standard Quality Report Structure). There are also other metadata/data reporting standards facilitating the access and reuse of public information, such as: i) Open archival information system (OAIS), specifies how to maintain, transfer and disseminate archival information across institutions, both metadata and data from public archives. The aim of this reference model is to acknowledge the actors, responsibilities/roles and procedures for the long-term maintenance of archival datasets considered public good [10]; and ii) Data Documentation Initiative (also known as DDI or DDI Metadata), an international standard only for metadata standardization in the case of micro data collected because of official statistics (surveys, questionnaires, etc.) conducted in National Statistics bodies [11].

This report is based on the results of the first WP8 task, with the aim to summarize existing knowledge and definition of health data, indicators, standardised data collection methods, availability and accessibility procedures covering different health data sources (e.g., population-based registries, surveys, longitudinal studies, health system

performance, other administrative sources, data collected through e-health solutions, etc.) across EU/EEA Member States (MSs).

In order to achieve the T8.1 goals, the WP8 team at the Italian National Institute of Health (ISS) developed a questionnaire to identify methods of data collection, and the related harmonization and standardization procedures, for health monitoring and HSPA in EU projects/studies. The structure of the questionnaire was decided upon the results of a scoping review of international organizations (i.e., WHO, Eurostat, OECD) and selected EU research networks (European Core Health Indicators Monitoring-ECHIM, Joint Assessment Framework on Health-JAF). The main findings of the scoping review were presented and discussed with WP8 partners during the first WP8 overall meeting held in Brussels (Belgium), in February 2019.

The survey on data collection methods and procedures across MSs has enabled the identification of national data collected for population health monitoring/public health surveillance and HSPA with standardized methods that are not incorporated into existing international datasets, such as WHO, OECD, or Eurostat. This survey is therefore an opportunity to identify and describe "isolated" projects/studies, which provide standardized and comparable health data at national or sub-national level. However, few projects/studies sharing health data with international organizations were still identified. The analysis of these projects, "isolated" or not, will be useful for further activities of WP8 that consist in developing guidelines for accessibility, availability and reporting of health information, including quality of health data, indicators and reporting.

## II. Methods

The first phase of T8.1 consisted of a scoping review of international organizations (i.e., WHO-Health For All database, WHO-Health 2020 monitoring framework, WHO-Global non-communicable diseases monitoring framework, OECD, Eurostat) and selected EU research networks (i.e., ECHIM, JAF) to identify HI data and metadata characteristics that could be used to develop a questionnaire on health data collection methods and the related harmonization and standardization procedures for health monitoring and HSPA across MSs. The identified data and metadata characteristics were then grouped into five main topics:

i) *Source of information*, types of data sources used (e.g., EHES, census, administrative data);
ii) *Methodology*, tools and approaches for data collection (e.g., questionnaires, face-to-face interviews, medical examination);
iii) *Quality control assessment*, quality assurance procedures and quality dimensions or criteria considered (e.g., relevance, clarity, comparability);
iv) *Availability* of micro or macrodata, data formats (e.g., digital, printed), and metadata standards;
v) *Accessibility*, standard for exchange and sharing of statistical data and metadata (e.g., request and approval required for data access; data are transferable to approved users and reusable; request for financial charge for data access).

In the second phase of T8.1, a questionnaire based on the aforementioned topics was developed to identify data collection methods, availability and accessibility of HI in projects/studies performed in Europe (Appendix 1). The projects/studies could be part of European research networks (e.g., EHES, ECHIM, European Collaboration for Healthcare Optimization-ECHO, European Cardiovascular Indicators Surveillance Set-EUROCISS, etc.) [12-15], but the related data or indicators are not included in databases of international organizations (e.g., WHO-Europe, OECD, Eurostat). Practical examples regarding health monitoring are: i) the Italian health examination survey [16] is included in EHES; ii) the Italian injury data is included in the European Injury Data Base (IDB) [16]; and iii) the Italian perinatal data is included in the European Perinatal Health Surveillance System (Euro-Peristat) [17]. An example for HSPA regards hospital-specific indicators from administrative databases and medical records in European countries, including Italy, that are currently being developed and tested as indicators of system performance (e.g. increased survival rates after acute cardiovascular events, including stroke and acute myocardial infarction [18].

The projects/studies eligible for the survey should satisfy all the following:

i) health data provided by the project/study should be representative of the population at national or regional level;
ii) health data should cover topical areas of population health monitoring and/or health system performance assessment;
iii) the project/study should not focus on rare diseases, infectious diseases and cancer;

iv) health data should be accessible as micro or macrodata (aggregated results) which are not included in databases of international organizations such as WHO, Eurostat, OECD;

v) the project/study produced scientific outputs (e.g. research papers, reports, etc.).

Regarding quality assurance procedures in data collection/data sources, the quality dimensions or criteria defined by Eurostat [19] (i.e., relevance, accuracy, timeliness, punctuality, comparability, coherence, accessibility and clarity) in addition to two quality criteria considered by ECHO (coverage and internal reliability) [14] were used in the questionnaire. Furthermore, the sections of the questionnaire on health data availability and accessibility were developed according to FAIR Data Principles, which are a set of guiding principles in order to make data FINDABLE (data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier); ACCESSIBLE (metadata and data are understandable to humans and machines, and data is deposited in a trusted repository); INTEROPERABLE (metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation); and REUSABLE (data and collections have a clear usage license and provide accurate information on provenance) [20].

The questionnaire was piloted, from April to May 2019, by administering the tool to the 25 participants of the first WP8 overall meeting. After editing the survey items according to the comments received, the final version of the questionnaire was administered to the representatives from InfAct partner countries (28 MSs and 4 associated countries) through the LimeSurvey online platform. The participants were asked to forward the questionnaire to their colleagues with good knowledge and experience in health monitoring/public health surveillance and HSPA in their country, such as epidemiologists, researchers that have played leading roles in EU projects, health data managers engaged in national health and research institutions, and universities (snowball recruitment). A set of definitions was provided to the participants, through an online page, to facilitate the comprehension of the survey items (Appendix 2). Data collection was carried out from June to October 2019.

The third phase of T8.1 consists of data analysis, reporting and dissemination of the research findings. The preliminary results of the survey were presented during the 12th European Public Health Conference 2019 held in Marseille, France. The survey results will be presented in September 2020 at the EuroScience Open Forum (ESOF) in Trieste (Italy). In addition, a research paper outlining the main findings of the study is under preparation.

# III. Results

## 1. General characteristics of the respondents

The results of the interim analysis are based on data related to 91 projects/studies from 18 EU countries (Belgium, Croatia, Czech Republic, Estonia, Finland, France, Germany, Italy, Latvia, Luxembourg, the Netherlands, Portugal, Romania, Serbia, Slovenia, Spain, Sweden, and the United Kingdom). As shown in Figure 1, the respondents of the survey are mostly affiliated with Public Health Institutes (46/91projects), Universities (23/91 projects) and Research Institutes (9/91 projects).

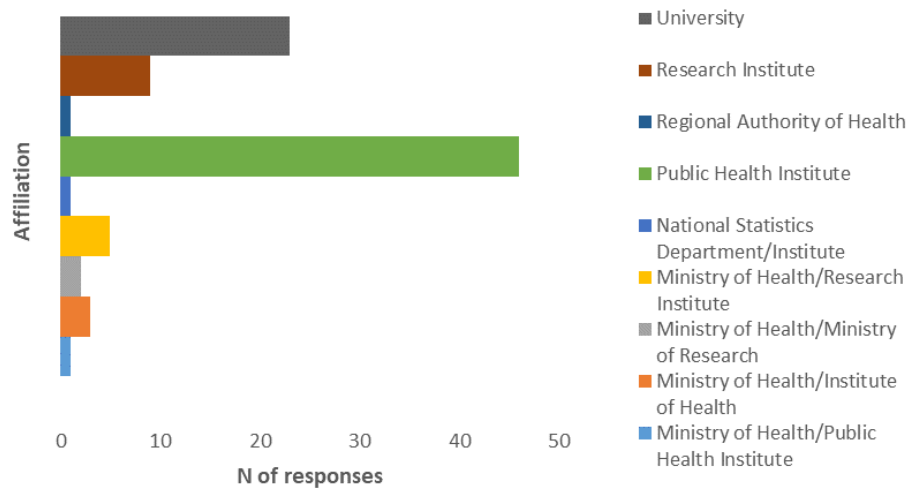


Figure 1. Affiliations of the survey respondents

## 2. Source of information/data sources

The 91 projects/studies (Appendix 3) are representative at national (45/91 projects), regional (20/91), or both national and regional levels (26/91). Some projects/studies are also research networks, for instance the Burden of Disease Network (BOD), European Perinatal Health Surveillance System (Euro-Peristat), European Health Examination Survey (EHES), Countrywide Integrated Noncommunicable Diseases Intervention (CINDI Network), Longitudinal study on Aging (ILSA), and more. The authorities or organizations responsible for the projects/studies (Figure 2) are mostly National Institutes of Public Health (25/91), National Health Institutes (17/91), and Universities (14/91).

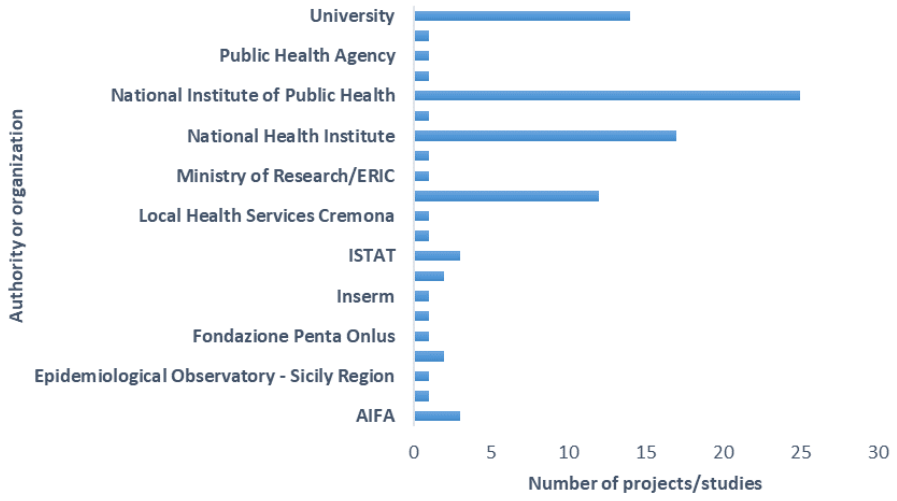


Figure 2. Authority or organization responsible for the projects/studies

The main objectives of the projects/studies (Table 1) are elaboration of health monitoring indicators (70/91), health data collection (57/91), elaboration of health system performance assessment indicators (30/91), and standardization and harmonization of methods and procedures (29/91). Other objectives were reported in 21 projects, for instance assessment of health outcomes related to environmental exposures, evaluation of safety/effectiveness/cost of pharmacological or oncological treatments and newly marketed drugs, data linkage, and more.

Table 1. Main objectives of the projects/studies

| Main objectives | N projects/studies |
|---|---|
| Health data collection | 57 |
| Elaboration of health monitoring indicators (e.g. prevalence, incidence, etc.) | 70 |
| Elaboration of health system performance assessment indicators (e.g. hospital-acquired infections, average length of stay, etc.) | 30 |
| Standardization and harmonization of methods and procedures | 29 |
| Development and/or validation of specific tools | 23 |
| Classifications and guiding principles | 8 |

| | |
|---|---|
| Other:<br>- Assessment of health outcomes related to environmental exposures<br>- Association between risk factors and chronic diseases (cardiovascular diseases)<br>- Data linkage<br>- Data visualization of Burden of Disease data for Germany<br>- Description of the situation with uninsured population and policy options<br>- Development of a prognostic score<br>- Elaboration of determinants of childhood obesity indicators<br>- Integrate multiple sources and data analysis<br>- Recommendations<br>- Safety assessment of pharmacological treatments, including adverse drug reactions (ADRs), evaluation of inappropriate prescribing, safety of newly marketed type 2 diabetes drugs<br>- Safety of oncological treatment for metastatic colorectal cancer<br>- Safety/effectiveness/costs of oncological treatments<br>- Scientific articles<br>- Surveillance system<br>- Target public health goals<br>- To identify new risk factors for chronic diseases and to evaluate their impact in public health<br>- To investigate genetic and environmental risk factors for cardiovascular and cancer diseases<br>- To provide information for planning prevention interventions | 21 |

The data sources used in the project/studies (Table 2) are mostly administrative data (e.g. hospital discharge records, mortality, pharmaceutical prescription, etc.) (52/91), followed by population health interview surveys - HIS (22/91), electronic medical/health records (20/91), medical records/clinical data registries (19/92) and population-based disease registries (18/91). The least used data sources were intermediate linked data sources, geographic information/geospatial data, and e-health solutions, and some official statistics or routine data. Intermediate linked data sources were used in five projects/studies: i) two Italian projects Longitudinal Study on Aging (ILSA) and the Safety of non-steroidal anti-inflammatory drugs (SOS); ii) the Secure Anonymized Information Linkage (SAIL system) project performed in the UK; iii) the Slovenian study on incidence and prevalence of diabetes; and iv) the Health insurance coverage study performed in Estonia, in which six national registries (e.g. Population Register, Social Security Information System, Working Register, etc.) were linked through a personal ID-code. Three studies used geographic information/geospatial data, namely the Atlas of Variations in Medical Practice in the Spanish National Health Service project (Atlas VPM project) and two German projects, AdiMon Indicator System and BURDEN 2020. E-health solutions (mhealth devices) were used in two projects/studies performed in Latvia (Health Care Monitoring Datalink) and in the UK (SAIL system).

The time period for data collection varied greatly (Table 2). It was mostly continuous for administrative data (26/52), medical records/clinical data registries (17/18), electronic medical/health records (14/20), population-based disease registries (13/17), hospital

based registries (10/15), and clinical quality registries (4/7). The data collection period was mainly periodic for HIS (15/22), primary data collected by direct examination (8/15) or through interview (8/14), longitudinal studies (7/11) and HES (7/13). Single implementation was reported for various data sources, except for population-based disease registries and intermediate linked data sources. The most used data sources specified by the respondents are reported in Appendix 4.

Table 2. Health data sources used in the projects/studies

| **Data sources** | **Collection period** | | | | |
|---|---|---|---|---|---|
| | N projects | Single | Periodic | Continuous | Periodic interval (years) |
| Population health examination survey | 13 | x | x, ◊ | x | 3-5 |
| Population health interview survey | 22 | x | x, ◊ | x | 1-7 |
| Population-based disease registries | 17 | | x | x, ◊ | 1-4 |
| Hospital based registries | 15 | x | x | x, ◊ | m; 1-5 |
| Clinical quality registries | 7 | x | x | x, ◊ | 5 |
| Medical records/clinical data registries | 18 | x | x | x, ◊ | 5 |
| e-health solutions (mhealth devices) | 2 | x | | x | |
| Longitudinal or cohort study | 11 | x | x, ◊ | x | 2-4 |
| Administrative data | 52 | x | x | x, ◊ | m; 3-5 |
| Electronic medical/health records | 20 | x | x | x, ◊ | 1-5 |
| Intermediate linked data sources | 5 | | x, ◊ | x | 1 |
| Primary data collected by direct examination | 15 | x | x, ◊ | x | 2-10 |
| Primary data collected through interview | 14 | x | x, ◊ | | 2-5 |
| Other:<br>- Geographic information/geospatial data<br>- Media data<br>- Official statistics (e.g., school entrance examinations, birth statistics, land use statistics)<br>- Routine data (e.g., health insurance quality reports) | 6 | x | x | x, ◊ | 4 |

Legend: x, implemented; ◊, most frequent collection period; m, monthly

## 3. Health data collection methods

The most frequently used tools or approaches for health data collection (Figure 3) are mandatory reporting from data providers (34/91), self-administered questionnaires (32/91), record linkage of various data sources (32/91), and electronic medical/health

records (30/91). Other tools or approaches mentioned by the respondents for few projects were 24-hour dietary recall (diary type) and media data.

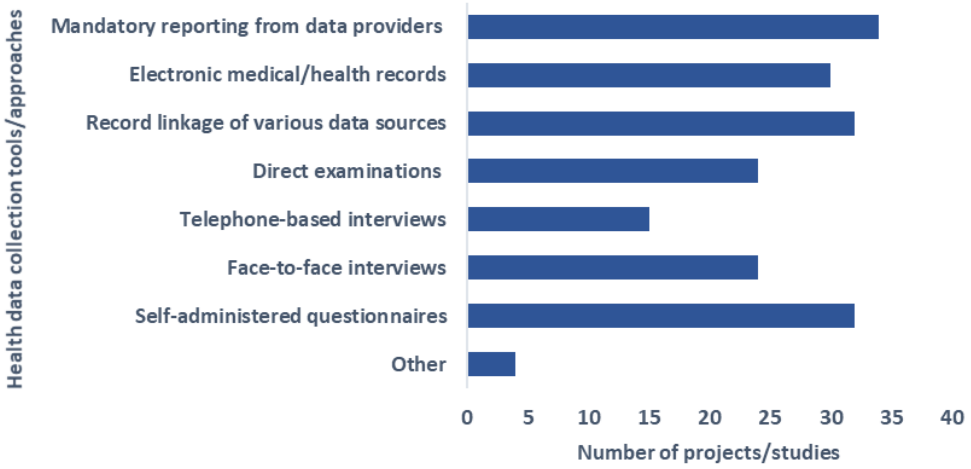


Figure 3. Tools or approaches used for health data collection

Health data collected or used by 30 projects/studies are shared with European research networks (e.g., ECHIM, ECHO, EHES, EHIS, EUBIROD, Euro-Peristat, Eurostat, etc.) (Table 3), while for 4 projects/studies the data sharing process is under development. Data sharing initiatives under development are between: i) the Italian project A plan for evaluating costs and outcomes of colorectal surgery in Emilia-Romagna (Emilia-Romagna Surgical Colorectal cancer Audit-ESCA) and the Dutch Colorectal Audit (DCRA) and the Dutch Institute for Clinical Audit (DICA); ii) different administrative registries in Finland and WHO, Eurostat, and OECD; iii) Czechs HES and EHES; and iv) Finland Health survey and EHES and ECHIM. The majority of projects/studies (57/91) do not share data with EU research networks.

Table 3. Projects/studies sharing data with European research networks

| Projects/studies | Research networks |
|---|---|
| Atlas VPM project | ECHO |
| Belgian Treatment Demand Indicator Register | EMCDDA |
| BURDEN 2020 | Eurostat |
| Drug-related mortality and hospitalization in Italy | Eurostat<br>EMCDDA<br>ECHIM<br>EU HFA database |
| Euro-Peristat | Euro-Peristat (coordinating center) |
| European Health Examination Survey | EHES (coordinating center) |
| European Injury Database | ECHIM, IDB |
| German Health Update - GEDA | Eurostat |
| HBSC | HBSC (coordinating center),<br>WHO |
| Health Interview Survey | EHIS (coordinating center) |
| Health Status Report | EHES (coordinating center) |
| Initiative for Quality Improvement and Epidemiology in Children and Adolescents with Diabetes | EUBIROD |
| Initiative for Quality improvement and Epidemiology in Diabetes | EUBIROD |
| Italian nationwide longitudinal population-based study on DKA at diagnosis of type 1 diabetes | Joint International Project DKA at onset of pediatric type-1 diabetes |
| Italian Obstetric Surveillance System | INOSS, Euro-Peristat<br>GBD Network |
| LINFA Project: Longitudinal Infant and Neonatal Follow-up towards Adolescence | EUROCAT<br>IARC<br>EURORDIS |
| Luxembourg's Birth-Related Health-Monitoring System - SUSANA | Euro-Peristat (coordinating center) |
| Luxembourgish Information System on Drugs and Drug Addiction | EMCDDA<br>REITOX |
| Moli-sani Study | MORGAM<br>Biomarcare consortium<br>CHANCES project |
| National Health Interview Survey | EHES<br>HBM4EU |
| Neonatal Hearing Screening | EUSCREEN |
| Observation of Cardiovascular risk factors in Luxembourg | NESCAV |
| OKkio alla SALUTE | WHO Europe-COSI |
| Secure Anonymized Information Linkage (SAIL) system | ECHIM, IDB |
| Survey of Health, Ageing and Retirement in Europe | SHARE-ERIC |
| CroDiab | ns |
| Romanian study | ns |

The projects/studies are related to health monitoring (84/91), health system performance monitoring (27/91) and health system performance assessment (21/91). They provide information mainly on non-communicable diseases (65/91), healthcare utilization (46/91), unhealthy lifestyles (35/91) and mental diseases (33/91) (Figure 4). Other conditions or health topics were also considered, such as infectious diseases, financial protection, frailty, illicit drug use, safety of pharmacological treatments, physical functioning, social functioning, ageing, social and geographical inequalities in health conditions, hearing and vision, oral health, judicial status, living conditions, maternal and child health, mortality, and sick-leave indicators.

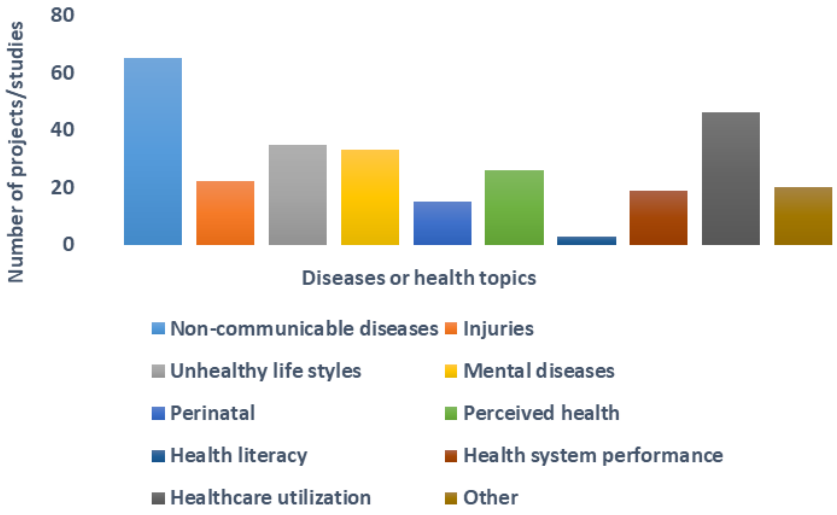


Figure 4. Main diseases or health topic of the projects/studies

The projects/studies also provided information on risk factors, high-risk conditions and/or health behaviours (Figure 5), mostly on socio-economic factors (47/91), diabetes (44/91), Body Mass Index - BMI (41/91), obesity (40/91 each) and hypertension (39/91). Other risk factors considered in 32 projects/studies were consumption of illicit substances, unsafe sex, causes and circumstances of injuries, cardiovascular events, sleep disorders, pharmacological treatments, respiratory health, and thyroid function.

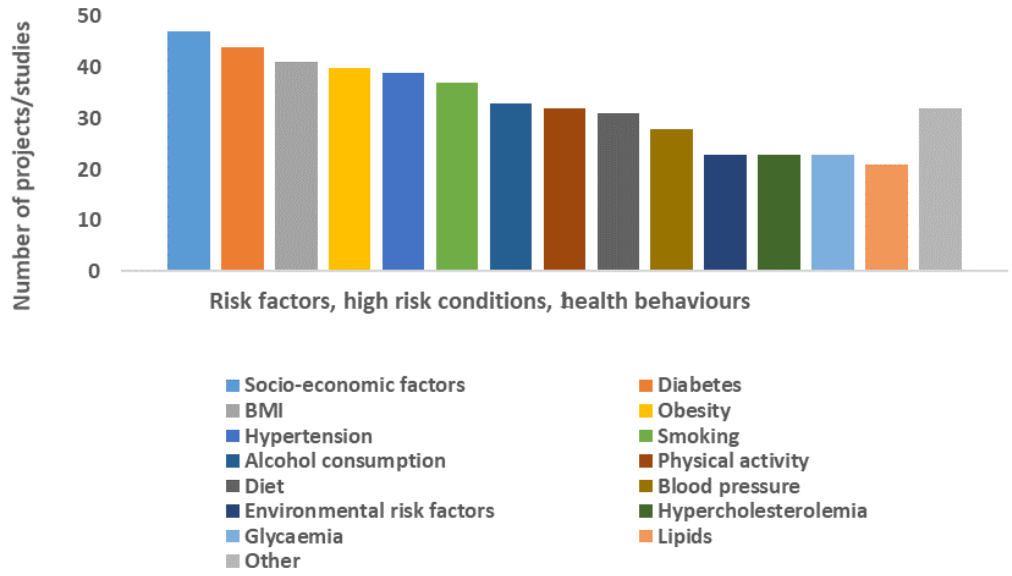


Figure 5. Risk factors, high risk conditions and health behaviours provided by the projects/studies

The areas defined in the protocols of the projects/studies (Table 4) are mostly related to statistical analysis (78/91), reporting (59/91) and quality data control (55/91). The protocols include internationally recognized standardized methods and procedures in all areas, but mostly for laboratory (17/17) and statistical analysis (50/78), quality data control (32/55), and reporting (32/59). For instance, standards and guidelines for laboratory analysis are provided by the Clinical and Laboratory Standards Institute (CLSI) while for biobanking, standardized methods are indicated by the European research infrastructure for biobanking (BBMRI-ERIC). Regarding statistical analysis and data quality control, the projects/studies adhere to guidelines and recommendations provided by international organizations (e.g., WHO, International Agency for Research on Cancer-IARC) and EU research networks (e.g., European Health Examination Survey-EHES, MONItoring of trends and determinants in CArdiovascular disease-MONICA, European Prospective Investigation into Cancer and Nutrition-EPIC). For reporting standards, the projects/studies follow the directives provided by international organizations (e.g., WHO, ECDC) , EU research networks (EHES, Health Behaviour in School-aged Children-HBSC, Infrastructure for Spatial Information in Europe-INSPIRE) and the recommendations of the International Committee of Medical Journal Editors (ICMJE) for the conduct, reporting, editing, and publication of research studies in medical journals.

Table 4. Areas defined in protocols of the projects/studies

| Areas of the protocol | N projects/studies | Internationally recognized standardized methods and procedures are reported in the protocol (N projects/studies) |
|---|---|---|
| Quality data control | 55 | 32 |
| Accessibility | 25 | 13 |
| Availability | 26 | 13 |
| Statistical analysis | 78 | 50 |
| Laboratory analysis | 17 | 17 |
| Reporting | 59 | 32 |
| Data linkage | 38 | 15 |
| Data sharing | 16 | 9 |

The indicators elaborated from the collected health data are mainly prevalence (59/91), outcome measures (52/91), incidence (47/91), performance measures (25/91), and attack rates (8/91). Other indicators mentioned by the survey respondents are epidemiological association measures (relative risk, odds ratio), case fatality rates, dietary habits, the number and profiles of patients under specialized drug treatment per year, prevalence and incidence estimates, mortality and hospitalization data, rates of events occurrence (deaths and discharges), evaluation of adverse drug reactions, sick-leave indicators, synthetic prognostic score, social and geographical inequalities indicators, temporal trends, and burden of disease indicators (disability-adjusted life years - DALYs; years lost due to disease - YLDs). The main uses of the elaborated indicators are monitoring (73/91), policy planning (66/91), research purposes (66/91) and health services evaluation (30/91). Other uses reported by some respondents are identification of trends and patterns in drug abuse, assessment of the use of health facilities, public health surveillance, and treatment.

The funding source for the majority of the projects/studies (84/91) is public (e.g., Ministry of Health, Ministry or Research, Italian Medicines Agency, European Food Safety Authority-EFSA). Other funding sources reported were pharmaceutical industries (i.e., Roche Pharma) and professional societies (i.e., Italian Society of Neurology).

## 4. Quality assurance procedures in health data collection

The most evaluated quality dimensions or criteria were relevance and comparability (65/88 each), followed by coverage (58/88), accuracy (52/88) and internal reliability

(47/88). Timeliness (38/88), coherence (35/88), and clarity (30/88) were also assessed in the projects/studies. The least evaluated criteria were punctuality and accessibility (28/88 each). Two projects/studies carried out in Estonia (the Health insurance coverage) and in Belgium (the Health status report) did not perform quality evaluation using the 10 quality criteria or dimensions indicated in the questionnaire (relevance, accuracy, timeliness, punctuality, comparability, coherence, accessibility, clarity, coverage and internal reliability). Quality assurance procedures were not reported for three projects/studies: Gruppo Italiano Reti Oncologiche- GIRO (Italy), the Treatment Demand Indicator Register (Belgium), and Euro-Peristat (France).

Regarding respondents' opinions about quality assessment of the health data, the results for the four most used data sources (administrative data, HIS, electronic medical/health records, and medical records/clinical data registries) are reported in Figures 6A-6D. Of 52 projects/studies using administrative data sources, the ten quality criteria were considered adequate or highly adequate in 35 or more projects/studies. The main criteria considered not adequate at all was accessibility regarding four projects/studies. Of 22 projects/studies using HIS, the quality criteria were considered adequate or highly adequate in 14 or more projects/studies. The only criteria considered not adequate at all was accessibility relatively to one project/study. In the 20 projects/studies using electronic medical/health records, the quality criteria were considered adequate or highly adequate in 6 or more projects/studies. None of the quality criteria was considered not adequate at all. Relevance, accuracy, comparability, coherence, accessibility and clarity were present but not adequate in 3 or more projects/studies. Of 19 projects/studies using medical records/clinical data registries, the quality criteria were considered adequate or highly adequate in 6 or more projects/studies. Punctuality, accessibility and clarity were considered not adequate at all for a maximum of four projects/studies.

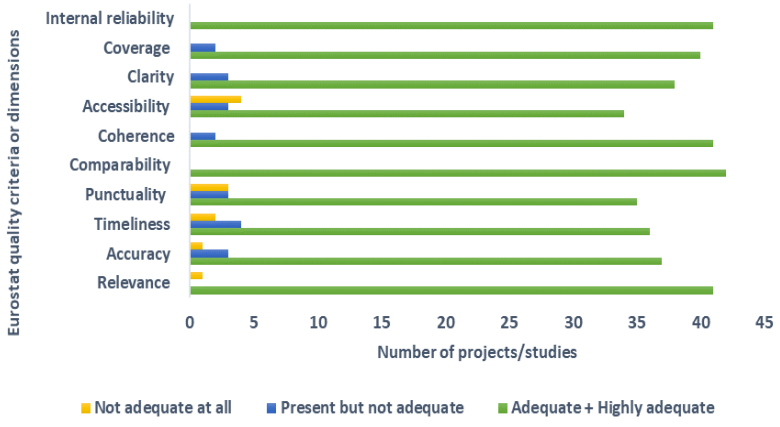

Figure 6A. Quality assessment of projects/studies using administrative data sources

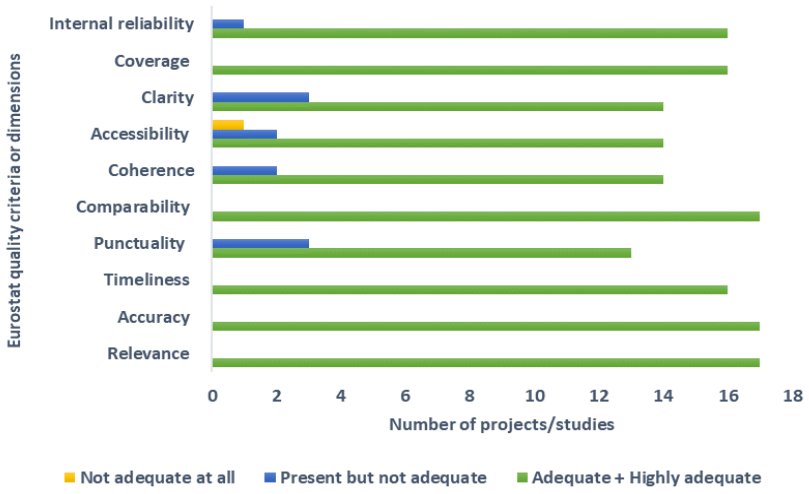


Figure 6B. Quality assessment of projects/studies using population health interview surveys

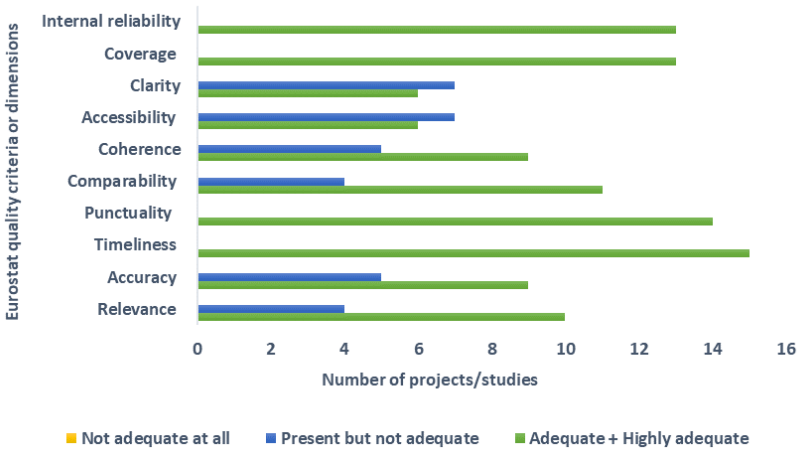


Figure 6C. Quality assessment of projects/studies using electronic medical/health records

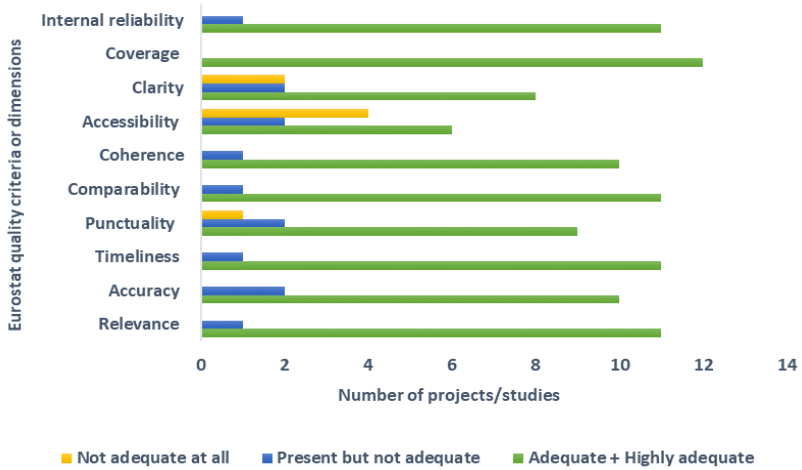

Figure 6D. Quality assessment of projects/studies using medical records/clinical data registries

## 5. Availability of health information

Collected health data are stored as microdata or individual records (41/86), macrodata or aggregated data (12/86), or both (33/86). Most projects/studies with microdata (59/74) have a global unique and eternally persistent identifier or study identifier. Of 44 projects/studies with macrodata, only 14 have an interactive system for users to perform further data aggregation and/or stratification. The available formats of the collected health data are first of all electronic files (75/86), followed by publications (40/86), websites (33/86) and CD-ROM in one project/study.

The majority of the projects/studies (50/84) has a publicly available description of the dataset purpose and content or metadata. The metadata follow reporting standards in 29 projects/studies, of which 7 are international reporting standards such as those defined by Eurostat (Appendix 2), 8 are national reporting standards, and 14 are ad-hoc metadata reporting standards developed for the purpose of a single project/study. Few international reporting standards were specified by the survey respondents. In particular, the Data Documentation Initiative (DDI) is used in the Finland Health survey and the SHARE project. DDI is an international standard for describing the data produced by surveys and other observational methods in the social, behavioural, economic, and health sciences [21]. The EU Injury Database (IDB) uses the Euro SDMX Metadata Structure (Appendix 2), as reported in the IDB Operating Manual [22]. Each IDB national metadata file reports information regarding the data quality (origin, content, methods of estimation, and other quality aspects) according to the principles of the European Statistical System [23] and the specifications of the IDB Operating Manual. Likewise, the German Health Update (GEDA) shares data with EHIS and the metadata to be collected under EHIS follow the European Statistical System standard specified by the Commission (Eurostat) [24]. Other projects/studies using international standardized protocols for metadata collection from EU participating countries are HBSC [25] and Euro-Peristat [26]. National and ad-hoc metadata reporting standards were not specified by the survey respondents.

## 6. Accessibility of health information

Health data are accessible to external users as macrodata (28/34) or microdata (21/34). Microdata are only available to users upon specific request followed by approval. The approval is mostly granted by a scientific committee (19/34) or through a formal agreement between institutions (17/34) (Figure 7A). Considering data reusability,

microdata are reusable based on data usage license (e.g., for a specific project, analysis, period of use, private or public use) in 26 projects/studies. They are reusable by all users in 4 projects/studies: 3 are conducted in Slovenia (the National Dietary Survey; Registry of sick-leave from work; Registry on Causes of Deaths) and the Survey of Health, Ageing and Retirement in Europe (SHARE) is performed in Luxembourg. Microdata are not reusable in the following 4 project/studies: i) Monitoring of health care quality indicators (Serbia); ii) Carte sanitaire (Luxembourg); iii) Health Care Monitoring Datalink (Latvia); and iv) the nationwide longitudinal population-based study on diabetic ketoacidosis (DKA) at diagnosis of type 1 diabetes (Italy).

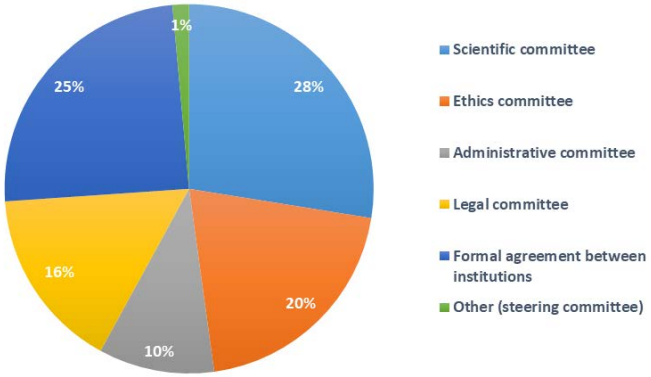


Figure 7A. Approval for microdata access

Macrodata are usually available to all users without specific request, thus in open access (22/22), or upon request followed by approval (18/22). The approval is mostly granted through a formal agreement between institutions (12/22) or by a scientific committee (9/22) (Figure 7B). Macrodata are reusable based on data usage license 22 projects/studies and are reusable for all users in 15 projects/studies. They are not reusable only in 2 projects/studies: Monitoring of health care quality indicators (Serbia) and Carte sanitaire (Luxembourg).

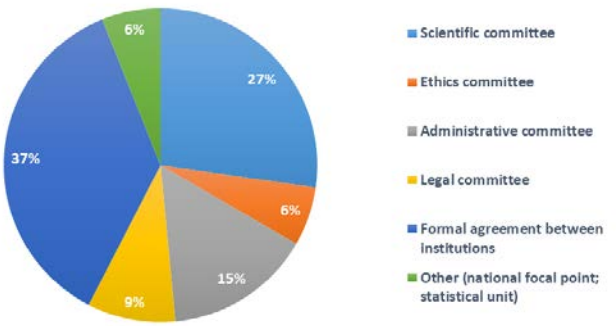

Figure 7B. Approval for macrodata access

The majority of the projects/studies do not provide a remote data access service for users (41/60), and a financial charge for data access is not required (44/60). The 16 projects/studies requiring a financial charge for data access are reported in Table 5.

Table 5. Projects/studies requesting a financial charge for data access

| | PROJECT/STUDY | COUNTRY |
|---|---|---|
| | German Health Update - GEDA | Germany |
| | Initiative for Quality Improvement and Epidemiology in Children and Adolescents with Diabetes | Belgium |
| | Initiative for Quality improvement and Epidemiology in Diabetes | Belgium |
| | Evaluation of ambulatory care quality | Belgium |
| | Initiative for Quality improvement and Epidemiology in multidisciplinary Diabetic Foot Clinics | Belgium |
| | Finland Health survey | Finland |
| | Different administrative registries | Finland |
| | FinSote | Finland |
| | German Health Interview and Examination Survey for Adults | Germany |
| | German Health Interview and Examination Survey for Children and Adolescents | Germany |
| | Nivel Primary Care Database | Netherlands |
| | Secure Anonymised Information Linkage (SAIL) system | United Kingdom |
| | Doetinchem Cohort Study | Netherlands |
| | Health Interview Survey | Belgium |
| | Italian nationwide longitudinal population-based study on DKA at diagnosis of type 1 diabetes | Italy |
| | Surveillance of cardiovascular diseases | France |

## IV. Implications for further research, challenges and conclusions

The survey findings highlight the heterogeneity in data collection methods and procedures and underline the lack of available, accessible, comparable or reusable health data and

information for research purposes and policy making in EU countries. In fact, only one-third of the identified projects share data with other EU research networks, one-sixth of the projects evaluate all 10 quality dimensions defined by Eurostat, and less than half of the projects follow metadata reporting standards for data description. The preliminary results of T8.3, Guidelines for accessibility and availability of health information, show similar findings. The ongoing review of EU research networks, under T8.3, is based on the five main topics identified through the scoping review of international organizations and research networks (i.e., source of information, methodology, quality control assessment, availability of health data and metadata, accessibility of health data and information) that were used for the development of the survey questionnaire. Up to date, 57 research networks collecting data on various health topics and representative at national or international levels have been identified. In these networks, data are mainly collected through administrative sources, health surveys and cohort studies. Less than half provide information on quality assessment of their data collection procedures and few networks share data with other research networks or specify the metadata-reporting standards used for data description. These findings underline the limits in health data usage and sharing within and across MSs.

The main challenge of the cross-sectional study was the correct identification and contact of survey respondents. InfAct participants were highly collaborative and assisted the ISS team in this task by forwarding the questionnaire to public health professionals engaged in health data management at national or international level. No doubt that this is a convenience sampling method but it enabled the distribution of the survey instrument in all 28 MSs and 4 associated countries.

Improving health information is a priority in Europe. The results of the survey will be used to identify examples of best practices in health data collection procedures, availability, and accessibility. A guidance document on data collection and data sharing methods will be developed based on the findings of the cross-sectional study and the ongoing review of EU research networks. The findings of the study on data collection methods and the ongoing review of research networks, as part of the InfAct web-based platform for health information research in EU, will facilitate the assessment of health inequalities across EU countries in terms of quality, availability, accessibility and comparability of health data and information. The research outputs will also facilitate sharing and dissemination of standardised and comparable health data collections, which are essential for research and evidence-based policy-making.

## References


1. Glasziou P, Irwig L, Mant D. Monitoring in chronic disease: a rational approach. BMJ 2005; 330(7492):644-8
2. Teutsch SM, Thacker SB. Planning a public health surveillance system. Epidemiol Bull. 1995; 16(1):1-6
3. Hurtado MP, Swift EK, Corrigan JM (eds). Envisioning the National Health Care Quality report. Washington: National Academies Press (US); 2001
4. Kelley E, Hurst J. Health care quality indicators project conceptual framework paper. Paris: Organisation for Economic Co-operation and Development (Health Working Papers no. 23); 2006
5. Wenzel H. Effectiveness, efficiency and equity. In: Kovacic L, Zaletel-Kragelj L (eds). Management in health care practice. A Handbook for teachers, researchers and health professionals. Hans Jacobs Publishing Company, Zagreb 2008
6. MacDonald PDM. Methods of field epidemiology. Jones & Bartlett Learning, Burlington, 2012
7. Eurostat: methodology. https://ec.europa.eu/eurostat/web/health/methodology
8. European Health Examination Survey: Guidelines for health examination surveys. http://www.ehes.info/manuals.htm#manual
9. Eurostat. ESS reference metadata reporting standards. https://ec.europa.eu/eurostat/data/metadata/metadata-structure
10. Open Archival Information System (OAIS). https://www.iso.org/standard/57284.html
11. Data Documentation Initiative. https://www.ddialliance.org/explore-documentation
12. European Health Examination Survey (EHES). http://www.ehes.info/
13. European Community Health Indicators Monitoring (ECHIM). http://www.mc3.sk/?projects=echim
14. European Collaboration for Healthcare Optimization (ECHO). http://echo-health.eu/
15. European Cardiovascular Indicators Surveillance Set (EUROCISS). http://www.cuore.iss.it/eurociss/en/project/project.asp
16. Osservatorio Epidemiologico Cardiovascolare: Health Examination Survey http://www.cuore.iss.it/eng/factors/oec.asp

17. The European Injury Data Base (IDB) https://ec.europa.eu/health/indicators_data/idb_en
18. Euro-Peristat country teams. https://www.europeristat.com/index.php/our-network/country-teams.html#it
19. Eurostat: Standard Quality Report. Methodological Documents, Working Group "Assessment of quality in statistics", Luxembourg, 2-3 October 2003
20. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship, 2016. Scientific Data, 3, 160018. doi:10.1038/sdata.2016.18
21. De Bruijne, M., & Amin, A. (2010): Questasy: Online survey data dissemination using DDI 3. IASSIST Quarterly, 33(1), 10-15. https://doi.org/10.29173/iq645
22. EuroSafe-European Association for Injury Prevention and Safety Promotion, EU-Injury Database: Operating Manual -version September2016, EuroSafe, Amsterdam: 2016
23. Eurostat (2011): European statistics code of practice for national and Community statistical authorities, September 2011
24. COMMISSION REGULATION (EU) No 141/2013of 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). Official Journal of the European Union, 2013
25. HBSC Survey methods. http://www.hbsc.org/methods/index.html
26. Euro-Peristat. https://www.europeristat.com/index.php

# APPENDIX 1. QUESTIONNAIRE FOR MEMBER STATES REGARDING HEALTH DATA COLLECTION METHODS AND PROCEDURES

**INVITATION E-MAIL**

Dear colleague,
the aim of this survey is to identify projects/studies which collect health data for population health monitoring (HM)/public health surveillance and health system performance assessment (HSPA) at national/regional level. The projects/studies could be part of European research networks, such as EUROCISS, EHES, ECHIM, EUBIROD, ECHO or purely national data collections not yet included in European research networks. There is no need to report data collections which are already part of existing international databases of WHO, OECD or EUROSTAT.

The questionnaire asks about data collection methods, quality assurance, and availability and accessibility of health information in your country.

The project/study eligible for the survey should satisfy all the following:

1. health data provided by the project/study should be representative of the population at national or regional level in your country
2. health data should cover topical areas of population HM and/or HSPA
3. the project/study should not focus on rare diseases, infectious diseases and cancer
4. health data should be accessible as micro or macrodata (aggregated results) which are not included in databases of international organizations, such as WHO, Eurostat, OECD
5. the project/study produced scientific outputs (e.g. research papers, reports, etc.).

Further information on the survey is available in the background documents:

*Introduction, Objectives and Inclusion Criteria*

*Glossary of Terms*

The survey is conducted by the Italian National Health Institute, Rome, in close collaboration with partners from the Netherlands, Germany, Slovenia, Finland and Spain.

We would appreciate if you could share information on more than one project/study, according to the typologies reported in the Introduction section (e.g., population-based registries, hospital-

based disease registries, clinical quality registries, health examination surveys, longitudinal studies, administrative healthcare data, e-health solutions, medical records, etc.).

THE QUESTIONNAIRE SHOULD BE FILLED-IN SEPARATELY FOR EACH SPECIFIC PROJECT/STUDY

**INFORMED CONSENT**
The participation in this survey is voluntary.

Your answers will be held in strict confidentiality and used only for the purposes of this study. Should you have any questions regarding the survey, please contact Luigi Palmieri at luigi.palmieri@iss.it or Brigid Unim at brigid.unim@iss.it.

The questionnaire can take about 25 minutes. You can stop the survey at any time and continue it later.

Do you want to participate in this survey?
Choose one of the following answers
□ YES
□ NO

## 1.GENERAL CHARACTERISTICS OF RESPONDENTS

1. Name of the country: _____ (dropdown menu on the online version)

2. Last name of the contact person: _____ First name: _____

3. Type of institute:
□ Public Health Institute □ Research Institute
□ University □ National Statistics Department/Institute
□ International Organization □ Other, please specify _____
□ Ministry of health/ Ministry of research

4. Work telephone number: _____ 5. E-mail: _____@_____

## 2. SOURCE OF INFORMATION/DATA SOURCES - PROJECT/STUDY BACKGROUND INFORMATION

1. What is the name of the project/study: _____
☐ If available, provide a link to the website: _____

2. Which authority/organization is responsible for this project/study:
_____

3. Who is the contact person for this project/study:
Name: _____
E-mail address: _____

4. The project/study is representative at:
□ regional level
□ national level
□ both

5. Which are the main objectives of the project/study (select all that applies)?
□ Health data collection
□ Elaboration of health monitoring indicators (e.g. prevalence, incidence, etc.)
□ Elaboration of health system performance assessment indicators (e.g. hospital-acquired infections, average length of stay, etc.)
□ Standardization and harmonization of methods and procedures
□ Development and/or validation of specific tools
□ Classifications and guiding principles
□ Other, please specify _____

6. What type of health data sources are used (select all that applies)?
□ Population health examination survey (HES)
□ Population health interview survey (HIS)
□ Population-based disease registries
□ Hospital based registries
□ Clinical quality registries
□ Medical record or clinical data registries
□ e-health solutions (mhealth devices)
□ Longitudinal or cohort study
□ Administrative data (e.g. hospital discharge records, mortality, pharmaceutical prescription, etc.)
□ Electronic medical/health records
□ Intermediate linked data sources
□ Primary data collected by direct examination (DA AGGIUNGERE)
□ Primary data collected through interview (DA AGGIUNGERE)
□ Other, please specify _____
☐ (available after each selected item in question 5). Please, specify the name of the health data source(s) and, if available, provide a link to the website

_____

Please, specify if the data collection is:
□ Continuous
□ Periodic; please specify the interval _____
□ Single implementation

7. What type of tools or approaches are used for the health data collection (select all that applies)?
□ Self-administered questionnaires
□ Face-to-face interviews
□ Telephone-based interviews
□ Direct examinations
□ Record linkage of various data sources
□ Electronic medical/health records

□ Mandatory reporting from data providers (i.e., administrative data collection)
□ Other, please specify _____

8. Are health data collected/used by the project/study shared with European research networks (e.g. EUROCISS, EHES, ECHIM, EUBIROD, ECHO, EuroREACH, etc.)?
□ Yes
□ No
□ Under development
☐ If yes, please specify the research network:
□ _____
□ _____
□ _____

☐ If under development, please specify the research network:
□ _____
□ _____
□ _____

9. How is the project/study funded (select all that applies)?
□ Public
□ Private
□ Other, please specify _____

10. Please specify if the project/study is related to:
□ Health monitoring
□ Health system performance monitoring
□ Health system performance assessment

11. On which of the following main diseases or health topics did the project/study provide information (select all that applies)?
□ Non-communicable diseases (e.g. cardiovascular, cancer, pulmonary, diabetes, etc.)
□ Injuries
□ Unhealthy lifestyles
□ Mental diseases
□ Perinatal
□ Rare diseases
□ Perceived health
□ Health literacy
□ Health system performance
□ Healthcare utilization
□ Other, please specify _____

12. On which of the following main risk factors, high-risk conditions and health behaviors did the project/study provide information (select all that applies)?
□ Blood pressure
□ Hypertension
□ Lipids

□ Hypercholesterolemia
□ Glycaemia
□ Diabetes
□ BMI
□ Obesity
□ Smoking
□ Alcohol consumption

□ Physical activity
□ Diet
□ Socio-economic factors
□ Environmental risk factors
□ Other, please specify _____

13. Which of the following areas is defined in the project/study protocol (select all that applies):
□ Quality data control
□ Accessibility
□ Availability
□ Analysis
□ Reporting
□ Data linkage
□ Data sharing
□ Other, please specify _____

☐ (after each selected item in question 13) Does the protocol include internationally recognized standardized methods and procedures for the selected areas?
□ Yes
□ No

☐ (available if "yes" for each selected item in question 13) Please, specify the reference or provide a link to the standardized methods and procedures
_____

14. Which are the main indicators elaborated from the collected health data (select all that applies)?
□ Prevalence
□ Incidence
□ Attack rates
□ Performance measures
□ Outcome measures
□ Other, please specify _____

15. What is the main use of the elaborated indicators (select all that applies)?
□ Monitoring
□ Policy planning
□ Research
□ Health services evaluation

□ Other, please specify _____

## 3. QUALITY ASSURANCE PROCEDURES IN DATA COLLECTION

The quality of statistical information is composed of the following dimensions or criteria: relevance, accuracy, timeliness and punctuality, comparability, coherence, accessibility and clarity (see glossary of terms).

1. Considering the above definition and the specific project/study indicated in section 2 - question 1, which dimensions or criteria are evaluated in quality assurance procedures at the national level (select all that applies)?

□ Relevance
□ Accuracy
□ Timeliness
□ Punctuality
□ Comparability
□ Coherence
□ Accessibility
□ Clarity
□ Coverage
□ Internal reliability
□ Other, please specify _____
□ All above mentioned dimensions or criteria
□ None of the above

2. For each of the selected data sources in section 2 - question 6, please provide your opinion/judgment regarding quality assessment of the health data in the table below

Data source 1 (only those data sources selected in section 2 - question 6 will be shown)

□ Population health survey (HES)
□ Population health survey (HIS)
□ Population-based disease registries
□ Hospital based registries
□ Clinical quality registries
□ Medical record or clinical data registries
□ e-health solutions (mhealth devices)
□ Longitudinal or cohort study
□ Administrative data (e.g. hospital discharge records, mortality, pharmaceutical prescription, etc.)
□ Electronic medical/health records
□ Intermediate linked data sources
□ Other, please specify _____

| Quality assessment criteria | Description* | Highly adequate | Adequate | Present but not adequate | Not adequate at all | Comments |
|---|---|---|---|---|---|---|
| Relevance | Degree to which statistics meet current and potential user needs | | | | | |
| Accuracy | Closeness of computations or estimates to the (unknown) exact or true values | | | | | |
| Timeliness | Length of time between its availability and the event or phenomenon it describes | | | | | |
| Punctuality | Time lag between the release date of data and the target date when it should have been delivered | | | | | |
| Comparability | Measure of the impact of differences between geographical areas, non-geographical domains, or over time | | | | | |
| Coherence | Adequacy to be reliably combined in different ways and for various uses | | | | | |
| Accessibility | Physical conditions under which users can obtain data | | | | | |
| Clarity | Availability of data information (documentation and metadata, illustrations, limitation in use, etc.) | | | | | |
| Coverage | The extent to which the sample stored describes actual performance. | | | | | |
| Internal reliability | A measure of whether the information stored is consistent over the years. | | | | | |

*See glossary of terms

- Data source n……. (If other data sources are indicated in section 2)

## 4. AVAILABILITY

1. Are the collected health data stored as micro (individual record) and/or macrodata (aggregated data)?

□ Microdata
□ Macrodata
□ Both

☐ If microdata are available, is there a global unique and eternally persistent identifier (study identifier)?
□ Yes
□ No

☐ If macrodata are available, is there an interactive system for users to perform further data aggregation and/or stratification?
□ Yes
□ No

2. Which are the available formats of the collected health data (select all that applies)?
□ Publication(s) (please specify the reference of relevant publication(s) _____)
□ Electronic files
□ CD-ROM
□ Websites (please specify the link _____)

3. Is there a publicly available description of the dataset purpose and content (metadata)?
□ Yes
□ No
☐ If yes, please provide a web-link(s) to the public information
_____

4. Do metadata follow reporting standards (e.g. SIMS, ESMS, ESMS-IP, ESQRS, OAIS, DDI described in Introduction section)?
□ Yes
□ No
□ I do not know/not aware

## 5. ACCESSIBILITY

1. Are the collected health data accessible to external users?
□ Yes, microdata
□ Yes, macrodata (aggregated data)
□ No

☐ If "yes microdata", the data are
□ available to users upon specific request followed by approval
□ available to all users without specific request (open access)

☐ If "yes macrodata (aggregated data)", the data are:
□ available to users upon specific request followed by approval
□ available to all users without specific request (open access)

If access is based on approval, how is the approval granted (select all that applies)?
□ By a scientific committee
□ By an ethics committee
□ Administrative committee
□ Legal committee

□ Formal agreement between institutions
□ Other (please specify): _____

2. Are data reusable (i.e. data have a clear usage licenses and provide accurate information on provenance)?
□ Yes, for all users
□ Yes, based on data usage license (e.g. for a specific project, analysis, period of use, private or public use)
□ No
If "yes, for all users", please specify if macro or microdata
If "yes, based on data usage license", please specify if macro or microdata

3. Is there a remote data access service provided for users?
□ Yes
□ No
☐ If yes, please provide the website address: _____

4. Is there a financial charge for data access?
□ Yes
□ No

*We thank you for your participation in this survey.*
*If you have another project/study to share with us, please click the following link: 'survey link'.*

# APPENDIX 2. GLOSSARY OF TERMS

- **Data:** characteristics or information, usually numerical, that are collected through observations [1]
- **Dataset:** any organized collection of data. The data set lists values for each of the variables and for each member of the dataset [2]
- **Microdata:** consist of sets of records containing information on individual respondents or business entities. To protect the anonymity of respondents (persons, organizations), the access to microdata is restricted [3]
- **Macrodata:** data derived from microdata by statistics on groups or aggregates, such as counts, means, or frequencies [4]
- **Metadata:** explanatory texts documenting statistical data and providing summary information on definitions of populations, objects, variables, the methodology and quality and the statistical production process in general. A distinction is generally made between structural and reference metadata [2]:
    - ***Structural metadata*** are used to identify, formally describe or retrieve statistical data, such as dimension names, variable names, dictionaries, dataset technical descriptions, dataset locations, keywords for finding data etc. For example, structural metadata refer to the titles of the variables and dimensions of statistical datasets, as well as the units employed, code lists (e.g. for territorial coding), data formats, potential value ranges, time dimensions, value ranges of flags, classifications used, etc.
    - ***Reference metadata*** (sometimes called explanatory metadata) describe the contents and the quality of the statistical data from a semantic point of view. They include explanatory texts on the context of the statistical data, methodologies for data collection and data aggregation as well as quality and dissemination characteristics
- **Metadata reporting standards:** the main reference metadata-reporting standards used by Eurostat [5]
    - SIMS (Single Integrated Metadata Structure)
    - ESMS (Euro SDMX Metadata Structure)
    - ESMS-IP (Euro SDMX Metadata Structure – Indicator Profile)
    - ESQRS (ESS Standard Quality Report Structure)

    There are also other metadata/data reporting standards facilitating the access and reuse of public information, such as:
    - Open archival information system (OAIS), specifies how to maintain, transfer and disseminate archival information across institutions, both metadata and data from public archives. The aim of this reference model is to acknowledge the actors, responsibilities/roles and procedures for the long-term maintenance of archival datasets considered public good [6]
    - Data Documentation Initiative (also known as DDI or DDI Metadata), an international standard only for metadata standardization in the case of micro data collected because of official statistics (surveys, questionnaires, etc.) conducted in National Statistics bodies [7].
- **Source of information/data sources:** specific datasets, metadata sets, databases or metadata repositories where data or metadata are available. According to the various ways in which data are collected, data sources can be distinguished in administrative, survey and registry sources [4]
- **Quality assurance procedures in data collection/data sources:** Eurostat [8] defines quality of statistical information in terms of the following dimensions or criteria: relevance, accuracy, timeliness and punctuality, comparability, coherence, accessibility and clarity

- ***Relevance*** is the degree to which statistics meet current and potential user needs. It refers to whether all statistics that are needed are produced and the extent to which concepts (definitions, classifications etc.) reflect user needs
- ***Accuracy*** in the general statistical sense denotes the closeness of computations or estimates to the (unknown) exact or true values
- ***Timeliness*** of information reflects the length of time between its availability and the event or phenomenon it describes
- ***Punctuality*** refers to the time lag between the release date of data and the target date when it should have been delivered, for instance, with reference to dates announced in some official release calendar, laid down by regulations or previously agreed among partners
- ***Comparability*** aims at measuring the impact of differences in applied statistical concepts and measurement tools/procedures when statistics are compared between geographical areas, non-geographical domains, or over time
- ***Coherence*** of statistics is their adequacy to be reliably combined in different ways and for various uses. When originating from different sources, and in particular from statistical surveys of different nature and/or frequencies, statistics may not be completely coherent in the sense that they may be based on different approaches, classifications and methodological standards
- ***Accessibility*** refers to the physical conditions under which users can obtain data: where to go, are access to data free or restrictive, etc.
- ***Clarity*** refers to the data's information environment whether data are accompanied with appropriate documentation and metadata, illustrations such as graphs and maps, whether information on their quality is also available (including limitation in use etc.) and the extent to which additional assistance is provided

- **Other quality dimensions or criteria considered by ECHO are** [9]:
  - ***Coverage:*** measures the extent to which the sample stored describes actual performance. Also represents a measure of the potential relevance of the data stored.
  - ***Internal reliability:*** a measure of whether the information stored is consistent over the years. It is a necessary condition for accurate estimations
- **Availability:** availability of micro or macro data, in various formats (publications, files, CD-ROM, Internet, etc.) and documentation related to various aspects of the data, such as methodological documents, summary notes or papers covering concepts, scope, classifications and statistical techniques [8, 10]
- **Remote data access service:** a service providing access to data stored on a computer or network from a remote distance. Remote data access services are often secured to ensure that users can only access data to which they have been approved and that users cannot alter or withdraw/copy the data from the system without permission [11]
- **Health Examination Survey (HES):** population based and objective surveys that provide data on many health indicators to support policy making, preventive activities and research. HES include questionnaire about socio-economic, demographic and health issues, as well as objective physical measurements, such as weight and blood pressure, and collection of biological samples, such as blood or urine [12]
- **Health Interview Survey (HIS):** collection of health status, healthcare use, health determinants and socio-economic background variables of a representative sample of the population living in private households through standardized questionnaires. The European Health Interview Survey (EHIS) includes information from all European Union Member States and is to be conducted every five years. EHIS is used as a data source for

important health and social policy indicators such as the European Core Health Indicators (ECHI) [13]

- **Population-based disease registry/register:** in epidemiology, the term register is applied to the file of data concerning all cases of a particular disease or other health-relevant condition in a defined population such that the cases can be related to a population base. With this information, incidence rates can be calculated. If the cases are regularly followed up, information on remission, exacerbation, prevalence, and survival can also be obtained. The register is the actual document and the registry is the system of ongoing registration [14]
- **Hospital-based disease registries** contain data on all patients with a specific type of disease diagnosed and treated at that hospital (e.g. cancer registries). There are two sub-categories under hospital-based registries: single hospital registries and multi-institution registries. The primary goal of the single hospital (institution) registry is to improve patient care by medical audit-type evaluation of outcomes [15]
- **Drug registries** (e.g. AIFA) record drugs and therapeutic plans submitted to monitoring [16]
- **Medical records** or **clinical data registries (e.g., Health search project, Pedianet project)** contain data on diagnoses, prescriptions and health assessments performed during each encounter with the patient and are recorded as part of the daily practice of physicians [17]
- **Clinical quality registries** (e.g. Sweet project, Pediatric Diabetes 2016) are organizations which systematically monitor the quality (appropriateness and effectiveness) of healthcare, within specific clinical domains, by routinely collecting, analyzing and reporting health-related information. They then feed this information back to clinicians to inform clinical practice and decision making [18]
- **Administrative source:** register of units and data associated with an administrative regulation (or group of regulations), viewed as a source of statistical data [2]
- **Survey:** investigation about the characteristics of a given population by means of collecting data from a sample of that population and estimating their characteristics through the systematic use of statistical methodology [2]
- **Longitudinal or cohort study:** observation of the population for a sufficient number of person-years to generate reliable incidence or mortality rates in the population subsets. This generally implies study of a large population, study for a prolonged period (years), or both [14]
- **e-health solutions:** e-health is an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a commitment to improve healthcare locally, regionally, and worldwide by using information and communication technologies [19]. Examples of e-health solutions are: i) electronic medical records or electronic health records; mobile health devices (mHealth) collecting survey data, ii) mobile payment processing technology to purchase fruits and vegetables; iii) EATFRESH.ORG is a healthy eating resource that offers multilingual information via its website, social media, and mobile technology; iv) Find MI Care is a free website and mobile application that simplifies the task of finding local, low-cost healthcare [20]

- **Healthcare performance measures**: measures that are commonly used to assess population health in relation to health-care performance. The measures focus on health insurance data as measure of occurrence, disease costs, or on patient data for quality assessment [21]
- **Indicator:** quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor [22]
- **Intermediate linked data source:** a database in which individual information from different sources are linked to contextual information (namely, demographic statistic, socioeconomic data and information on supply) to produce intermediate outputs or data that can further elaborated [9]

**REFERENCES**


1. OECD glossary on statistical terms. https://stats.oecd.org/glossary/detail.asp?ID=532
2. Eurostat glossary. http://ec.europa.eu/eurostat/statisticsexplained/index.php/Category:Statistical_concept
3. OECD Glossary of statistical terms. https://stats.oecd.org/glossary/detail.asp?ID=1574
4. OECD: glossary of statistics terms. htpps://stats.oecd.org
5. EUROSTAT. ESS reference metadata reporting standards. https://ec.europa.eu/eurostat/data/metadata/metadata-structure
6. Open Archival Information System (OAIS). https://www.iso.org/standard/57284.html
7. Data Documentation Initiative. https://www.ddialliance.org/explore-documentation
8. Eurostat: Standard Quality Report. Methodological Documents, Working Group "Assessment of quality in statistics", Luxembourg, 2-3 October 2003
9. European Collaboration for Healthcare Optimization (ECHO) Project. Zaragoza (Spain): Instituto Aragonés de Ciencias de la Salud - Instituto Investigación Sanitaria Aragón; c2010. Estupiñán Romero FR, Baixauli Pérez C, Bernal-Delgado E on behalf of the ECHO consortium. Handbook on methodology: ECHO information system quality report; 2014
10. SDMX Guidelines. SDMX glossary version 2.0, October 2018. https://sdmx.org/
11. OECD Health Statistics 2018: Frequently Requested Data (Excel file). Glossary of terms. http://www.oecd.org/health/publicationsdocuments/datasets/
12. European Health Examination Survey: National health examination surveys in Europe. http://www.ehes.info/national_hes.htm
13. Eurostat glossary: European health interview survey (EHIS). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:European_health_interview_survey_(EHIS)
14. Porta M, Greenland S, Last JM (eds): A Dictionary of Epidemiology. Oxford University Press, Oxford, Fifth edition, 2008

15. National Cancer Institute. SEER training Modules: Hospital-Based Registries. https://training.seer.cancer.gov/registration/types/hospital.html
16. Montilla S, Xoxi E, Russo P, Cicchetti A, Pani L. Monitoring registries at italian medicines agency: fostering access, guaranteeing sustainability. Int J Technol Assess Health Care 2015;31(4):210-3
17. National Quality Registry Network: What is a Clinical Data Registry? https://www.abms.org/media/1358/what-is-a-clinical-data-registry.pdf
18. Australian commission on safety and quality in healthcare. https://www.safetyandquality.gov.au/our-work/information-strategy/clinical-quality-registries/
19. Eysenbach G. What is e-health? J Med Internet Res. 2001; 3(2): e20
20. The Promises and Perils of Digital Strategies in Achieving Health Equity: Workshop Summary. Roundtable on the Promotion of Health Equity and the Elimination of Health Disparities; Board on Population Health and Public Health Practice; Health and Medicine Division; National Academies of Sciences, Engineering, and Medicine. Washington (DC): National Academies Press (US); 2016
21. Smith PC, Mossialos E, Papanicolas I, Leatherman S. Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects. WHO Europe/European Observatory on Health Systems and Policies, 2009
22. Glossary of the main terms used in results based management. www.oecd.org/development/peer-reviews/17484948.pdf

# APPENDIX 3. PROJECTS/STUDIES IDENTIFIED THROUGH THE SURVEY ON DATA COLLECTION METHODS AND PROCEDURES

| COUNTRY | PROJECT/STUDY | RN |
|---|---|---|
| Belgium | Belgian Treatment Demand Indicator Register | |
| Belgium | European Health Examination Survey | x |
| Belgium | Evaluation of ambulatory care quality | |
| Belgium | Health Interview Survey | x |
| Belgium | Health Status Report | |
| Belgium | Initiative for Quality Improvement and Epidemiology in Children and Adolescents with Diabetes | |
| Belgium | Initiative for Quality improvement and Epidemiology in Diabetes | |
| Belgium | Initiative for Quality improvement and Epidemiology in multidisciplinary Diabetic Foot Clinics | |
| Croatia | CroDiab | |
| Czech Republic | European Health Examination Survey | x |
| Estonia | The health insurance coverage study | |
| Finland | Different administrative registries | |
| Finland | Finland Health survey | |
| Finland | FinSote | |
| France | Euro-Peristat | x |
| France | Surveillance of cardiovascular diseases | |
| Germany | AdiMon Indicator System | |
| Germany | BURDEN 2020 | x |
| Germany | Health Interview and Examination Survey for Adults | x |
| Germany | Health Interview and Examination Survey for Children and Adolescents | |
| Germany | Health Update - GEDA | |
| Germany | National diabetes surveillance | |
| Italy | A plan for Evaluating Costs and Outcomes of colorectal Surgery in Emilia-Romagna (Emilia-Romagna Surgical Colorectal cancer Audit-ESCA) | |
| Italy | CAMUNI cerebrovascular disease registry | |
| Italy | CAMUNI Registry of Myocardial Infarctions | |
| Italy | CARENET - Performance evaluation and value assessment for cardiovascular and oncological care path in a regional network context: challenges and opportunities | x |
| Italy | COACH - Comparing Outcomes of Acute Cerebrovascular and other neurological Hospitalizations | |
| Italy | Developing and validating a new population-based risk stratification tool for predicting mortality, hospital admissions and healthcare costs | |
| Italy | Developing and validating a novel multisource comorbidity score from administrative data: a large population-based cohort study from Italy. | |
| Italy | Drug-related mortality and hospitalization in Italy | |
| Italy | Epidemiological Surveillance | |

| | | |
|---|---|---|
| Italy | EU-ADR - Exploring and understanding adverse Drug reactions by integrative mining of clinical records and biomedical knowledge | |
| Italy | European Injury Database | x |
| Italy | FABIO - Valutazione dell'utilizzo di FArmaci BIOlogici nel pazienti Oncologico | |
| Italy | FRAME - Flussi Regionali Automatizzati per il Monitoraggio dell'assistenza e la generazione di Evidenze scientifiche di indirizzo per le politiche sanitarie | x |
| Italy | GIRO - Gruppo Italiano Reti Oncologiche | x |
| Italy | GRETA - Generating Real-world Evidence on the Treatment of metastatic colorectal cancer with Avastin-bevacizumab | |
| Italy | Health Behaviour in School-aged Children (HBSC) | x |
| Italy | Improving microbiology diagnostic system quality in the function of surveillance on communicable diseases in the Republic of Serbia | |
| Italy | IST-02566 Differenze di mortalità e di ospedalizzazione secondo lo stato di salute, gli stili di vita e il consumo di servizi sanitari | |
| Italy | Italian Longitudinal Study on Aging - ILSA | x |
| Italy | Italian nationwide longitudinal population-based study on DKA at diagnosis of type 1 diabetes | x |
| Italy | Italian Obstetric Surveillance System (ItOSS) | x |
| Italy | Italian PRoject on the Epidemiology of Alzheimer's disease - IPREA | x |
| Italy | LINFA Project: Longitudinal Infant and Neonatal Follow-up towards Adolescence | |
| Italy | MACHINE - Mother And CHild-INfant real-world Experience | |
| Italy | Moli-sani Study | |
| Italy | MONICA-Brianza | x |
| Italy | National Registry of Major Coronary and Cerebrovascular Events | |
| Italy | Surveillance system OKkio alla SALUTE | |
| Italy | Surveillance system Passi d'argento | |
| Italy | Patterns of multimorbidity | |
| Italy | Pharmacological treatment in the elderly patient affected by cardiovascular disease and other chronic comorbidities: inappropriate prescribing and outcome evaluation among institutionalized and community-dwelling elders | |
| Italy | Population-based specialized gastric cancer registry in the province of Cremona | |
| Italy | Profili di salute | |
| Italy | Progetto PDTA - Metodologia per il monitoraggio e la valutazione dei percorsi diagnostico-terapeutico assistenziali (PDTA) nell'ambito del Nuovo Sistema di Garanzia dell'assistenza sanitaria | |
| Italy | QUADIM - I percorsi di cura nei disturbi mentali gravi, tra valutazione della qualità della cura e nuovi modelli di finanziamento | |
| Italy | Risk of Cardiovascular diseases and abdominal aortic Aneurysm in Varese (RoCAV) | |
| Italy | SAFEGUARD – Safety Evaluation of Adverse Reactions in Diabetes | |
| Italy | Socio-economic inequalities in mortality | |
| Italy | SOS - Safety of non-steroidal anti-inflammatory drugs | |
| Italy | Surveillance system PASSI | |
| Italy | The Viadana study | |
| Italy | Valutazione degli eventi avversi (cardio e cerebrovascolari) negli utilizzatori di incretine e altri antidiabetici attraverso l'analisi dei database amministrativi della Regione Lombardia | |
| Italy | Web-based antimicrobial surveillance tool | |

| | | |
|---|---|---|
| Latvia | Health Care Monitoring Datalink | |
| Luxembourg | Carte sanitaire | |
| Luxembourg | European Health Examination Survey | x |
| Luxembourg | European Injury Database | x |
| Luxembourg | Health Behaviour in School-aged Children (HBSC) | x |
| Luxembourg | Luxembourg's Birth-Related Health-Monitoring System - SUSANA | |
| Luxembourg | Luxembourgish Information System on Drugs and Drug Addiction | |
| Luxembourg | Neonatal Hearing Screening | |
| Luxembourg | Observation of Cardiovascular risk factors in Luxembourg - ORISCAV-LUX 1 &2 | |
| Luxembourg | SHARE - Survey of Health, Ageing and Retirement in Europe | x |
| Luxembourg | Study of infant feeding practices for babies aged 4, 6 and 12 months in Luxembourg | |
| Netherlands | Doetinchem Cohort Study | |
| Netherlands | Nivel Primary Care Database | |
| Portugal | National Health Interview Survey | x |
| Romania | Romanian study | |
| Serbia | Monitoring of health care quality indicators | |
| Slovenia | CINDI Health Monitor Survey | x |
| Slovenia | National Dietary Survey (EU-MENU) | |
| Slovenia | National Survey on Oral Health | |
| Slovenia | Registry of sick-leave (from work) | |
| Slovenia | Registry on Causes of Deaths | |
| Slovenia | Study on incidence and prevalence of diabetes | |
| Slovenia | Survey on use of alcohol, tobacco and illicit drugs | |
| Spain | Atlas of Variations in Medical Practice in the Spanish National Health Service (Atlas VPM project) | |
| Sweden | The National Public Health Survey | |
| United Kingdom | Secure Anonymized Information Linkage (SAIL) system | |

RN, Research Network

# APPENDIX 4. DATA SOURCES SPECIFIED BY THE SURVEY RESPONDENTS

| ADMINISTRATIVE DATA SOURCES* |
|---|
| National Statistical Office |
| Information system for health care data (data transparency) |
| Diagnosis-Related Groups Statistics |
| Assisted registry DB (anagraphical and demographic data) |
| Medical records from specialist services and general practice service |
| National Population Registry |
| Birth registry |
| Death registry |
| Outpatient procedures DB (outpatient visits, including visits in specialist ambulatories and diagnostic laboratories accredited by the NHS) |
| Abortions and spontaneous abortions informative flow |
| DB of liberal midwives |
| Emergency admission DB |
| Hospital discharge records |
| Registry on sick-leave (from work) |
| Pharmaceutical prescriptions, drug sales DB |
| Social Security database |
| Co-payment exception DB (co-payment exception for diagnosed chronic diseases) |
| Health insurance claims |
| Medical devices DB |
| **POPULATION HEALTH INTERVIEW SURVEY** |
| CINDI Health Monitor Survey (EU countries) |
| GEDA (Germany) |
| HBSC (Luxembourg) |
| Health Interview Survey (Belgium) |
| Health Interview Survey (Italy) |
| National Dietary Survey (Slovenia) |
| National Survey for Wales (UK) |
| ORISCAV study (Luxembourg) |
| Self-administered questionnaire on babies aged 4, 6 or 12 months (Luxembourg) |
| Survey on oral health (Slovenia) |
| Survey on use of alcohol, tobacco and illicit drugs (Slovenia) |
| The National Public Health Survey (Sweden) |
| **ELECTRONIC MEDICAL/HEALTH RECORDS** |
| Clinical charts (Italy) |
| Data collected in emergency departments from injury cases (Luxembourg) |
| EHRs from healthcare professionals (e.g., primary care physicians, other medical specialists) (the Netherlands, UK) |

| |
|---|
| **Health insurance fund (Estonia)** |
| **Hospital EHRs (including discharged records) (various countries)** |
| **Required immunization record (Serbia)** |
| **Medical record or clinical data registries** |
| **Clinical charts (paper or electronic) (Italy)** |
| **Clinical registry for diabetes (Italy)** |
| **Data collected in emergency departments from injury cases (Luxembourg)** |
| **Genetic counseling (Italy)** |
| **GP network (Belgium)** |
| **Health insurance claims data (Germany)** |
| **Hospital discharge records (Italy)** |
| **Birth registries (France)** |
| **Population-based disease registries** |
| **Cancer registry of Trento; Congenital Anomalies Registry of Trento/Mantova; Rare Diseases Registry of Trento (Italy)** |
| **Registries for diabetes, ischemic heart disease, cancer, tuberculosis (Belgium)** |
| **Register of congenital malformations; register of induced abortions; register of child welfare; register of primary health care visits; register of social assistance; register of sterilizations (Finland)** |
| **DPV registry for type 1 diabetes (Germany)** |
| **General Practice and hospital data registries (UK)** |
| **Mortality database (Italy)** |
| **Hospital Discharge Records (Italy, UK)** |
| **National Cancer Registry (the Netherlands)** |
| **Register of Patients with Rare Diseases (Latvia)** |
| **Registry of Type 1 Diabetes Mellitus-RIDI (Italy)** |

* data sources specified by respondents from the majority of the countries

Sciensano | Rue Juliette Wytsmanstraat 14 |

1050 Brussels | Belgium| e mail: infact.coordination@sciensano.be|