



PLATFORM FOR POPULATION-BASED REGISTRIES

Manual of operations

WP8 - task 1 report

30.10.2017

Project Leader:
Simona Giampaoli
Istituto Superiore di Sanità
Viale Regina Elena 299
00161 Rome, Italy
Tel +390649904231
Fax +390649904227
E-mail simona.giampaoli@iss.it



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

CONTENTS

EXECUTIVE SUMMARY	3
ABBREVIATIONS AND ACRONYMS	5
1. INTRODUCTION AND RATIONALE	7
1.1 The burden of non-communicable diseases.....	10
2. OBJECTIVES	12
3. POPULATION-BASED REGISTRIES	13
3.1 Definition and characteristics	13
3.2 Historical background of population-based registries.....	16
3.2.1 Cardiovascular diseases	17
3.2.2 Cancer	24
3.2.3 Rare diseases	26
3.2.4 Injury.....	27
3.2.5 Type 1 diabetes mellitus	31
3.2.6 Congenital hypothyroidism.....	34
3.2.7 Twins.....	35
3.2.8 Maternal mortality	36
3.2.9 Implanted prostheses.....	40
3.2.10 National Italian Transplant Centre	42
3.2.11 PARENT	42
4. SURVEILLANCE.....	43
4.1 Data sources for population-based registries.....	45
4.1.1 Death certificates	46
4.1.2 Hospital discharge and medical records	46
4.1.3 Autopsy registry	47
4.1.4 Nursing home and clinics.....	47
4.1.5 Emergency and ambulance services	48
4.1.6 General practitioner medical records.....	48
4.1.7 Drug dispensing registries.....	48
4.1.8 Other sources	48
5. METHODS	48
5.1 Planning a registry.....	49
5.1.1 Formulate the purpose of the registry	49

5.1.2	Determine whether the registry is the appropriate tool to achieve the purpose	50
5.1.3	Identify key stakeholders	51
5.1.4	Assess the feasibility of a registry	51
5.1.5	Build the team	52
5.1.6	Establish a governance	53
5.1.7	Define the needed scope and rigor	54
5.1.8	Define the data set, events, target population, indicators	55
5.1.9	Develop a protocol	60
5.1.10	Develop a project plan	60
6.	IMPLEMENTATION	61
6.1	STEP 1: Define target population and routine data	61
6.2	STEP 2: Carry out record linkage of administrative data	64
6.3	STEP 3: Perform a pilot study and validate routine data	65
6.4	STEP 4: Set up a population-based registry	66
6.5	STEP 5: Analyses and dissemination of results	67
7.	QUALITY CONTROL	68
7.1	Completeness	68
7.1.1	Completeness of event identification	69
7.1.2	Completeness of information	70
7.1.3	Evaluation of completeness	70
7.1.4	Qualitative method evaluation	70
7.1.5	Quantitative method evaluation	72
7.2	Validity	73
7.2.1	Internal validity	74
7.2.2	External validity	75
7.3	Assurance of data quality	77
8.	ETHICAL ISSUES	81
8.1	Fundamental principles	82
8.2	European regulation and data protection	85
8.2.1	Glossary	86
9.	ECONOMIC CONSIDERATIONS	94
10.	CONCLUSIONS	95
11.	THE NETWORK OF EXPERTS	96
12.	REFERENCES	98

EXECUTIVE SUMMARY

Population-based registries are useful in the monitoring of population health trends; they relate to a defined general population at risk, and allow calculating incidence rates. These registries can provide detailed information about various conditions, including personal characteristics, socio-economic conditions, risk factors, physical/biological measurements, lifestyles, stage of the disease, histology and survival. Hence, they are widely used in epidemiological research and also help assess the level of care performance.

The quality of population-based registry data is evaluated on the basis of data completeness, validity and timeliness.

Although registries are extremely useful for research, considerable resources are required for their implementation and maintenance.

This report is the result of a long and fruitful cooperation among experts, including epidemiologists, statisticians, clinicians and public health professionals. It provides simple recommendations to support and stimulate the implementation of population-based registers aimed at producing reliable and comparable estimates of disease occurrence indicators monitoring temporal trends and geographical gradients. The application of the recommended standard methodology in different countries will result in the availability of reliable, valid and comparable data on disease occurrence at European level and will facilitate the implementation of preventive actions.

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.

The added value of this report is:

- The creation of a network of experts to support the monitoring of population health trends across Europe;
- The offer of a step-wise procedure to implement disease occurrence indicators such as incidence and survival rates and case fatality (recommended European Core Health Indicators Monitoring);
- The establishment of a basis to improve future regulations of public health policies concerning surveillance across European countries.

Key points

- Population-based registries aim to identify all disease events occurred in a defined population
- The basic role of a population-based registry is to calculate incidence and survival rates, but can also provide other extensive information
- They are widely used in epidemiological research, to monitor time trends and geographical gradients, and set up preventive action programmes
- They are becoming more widely involved in clinical care processes
- Resources are required for their implementation and maintenance

ABBREVIATIONS AND ACRONYMS

ACC	American College of Cardiology
AHA	American Heart Association
AMI	Acute myocardial infarction
CH	Congenital hypothyroidism
CVD	Cardiovascular disease
DALY	Disability-adjusted life year
DRG	Diagnosis related group
DGSANTE'	
DGSANCO	Directorate-General for Health and Food Safety of the European Commission
EC	European Commission
ECHIM	European Core Health Indicators Monitoring
EHLASS	European Home and Leisure Accident Surveillance System
EMA	European Medicines Agency
ENRC	European Network of Cancer Registries
ESC	European Society of Cardiology
EU	European Union
EUROCISS	European Cardiovascular Indicators Surveillance Set
Eurostat	Statistical Office of the European Community
GDPR	General Data Protection Regulation
GP	General Practitioner
HDR	Hospital discharge record
HES	Health examination survey
HI	Health information
HIS	Health interview survey
IARC	International Agency for Research on Cancer
ICD	International Classification of Diseases
ICOR	International Consortium of Orthopaedic Registries
IDB	European Injury Database
IHD	Ischemic heart disease
INRICH	Italian National Registry of Infants with Congenital Hypothyroidism
IT	Information technology
ITR	Italian Twin Register
MDS	Minimum Data Set
MM	Maternal Mortality
MONICA	MONItoring Trends and Determinants in CARdiovascular Disease
MS	Member States
MTOS	Major Trauma Outcome Study

NCD	Non-communicable diseases
OECD	Organisation for Economic Co-operation and Development
PARENT	Patient Registries Initiative
PHP	Public Health Programme
PIN	Personal identification number
PPV	Positive predictive value
RD	Rare diseases
RIAP	Italian Arthroplasty Registry Project
RIDI	Italian Registry of Insulin-Dependent Diabetes Mellitus
RoR	Registry of registries
SHAR	Swedish Hip Arthroplasty Registry
WHO	World Health Organization
WHO-EUR	World Health Organization-Region of Europe

1. INTRODUCTION AND RATIONALE

Non-communicable diseases (NCDs) represent a huge public health problem in Europe, which results in a pressing need to implement comprehensive strategies to monitor and address this growing epidemic. Surveillance remains a primary need to evaluate the burden of disease, support research, plan preventive actions, assess public health outcomes, and influence decision-making policies. Although NCDs have been identified as one of the leading contributors to the global disease burden, the number of reliable and comparable indicators for which NCD data are available across Europe is currently limited.

International organizations, such as the European Commission (EC), the Organization for Economic Co-operation and Development (OECD), the World Health Organization-Regional Office for Europe (WHO-EUR), regularly publish data for groups of diseases with codes and classifications that are not always comparable. Moreover, they collect specific indicators by ad hoc surveys to respond to specific questions usually related to the different mission of each organization; they may refer to centres or research institutions directly linked to them (e.g. WHO Centres) that collect data without national standard methods or in small population samples. For these reasons, sometimes data reported do not correspond to national statistics [1,2].

Public health surveillance has been defined as *‘the continuous, systematic collection, analysis, interpretation and dissemination of data regarding health-related events for the planning, implementation and evaluation of public health actions to reduce mortality and morbidity and to improve health’* [3].

European Union (EU) projects previously supported by the health monitoring programme of the Directorate General for Health and Food Safety of the European Commission (today DG SANTE) in domains of population health and healthcare performance have developed methods to collect and process standardized data, assess indicators and deliver manuals of operations for health examination surveys (HES), establish protocols for human bio monitoring, injury and disease population-based registries, clinical and administrative health data collection systems and methods of health system monitoring and evaluation.

The BRIDGE Health - bridging information and data generation for evidence-based health policy and research [4], funded by the EC, started in 2015 with the following aims:

1. To enhance the transferability of health information (HI) and data for policy making and improve the utility and use of data and indicators for stakeholders in policy making, public health surveillance and healthcare;

2. To reduce HI inequality within the EU and within member states (MS);
3. To develop a blueprint for a sustainable and integrated EU HI system by developing common methods for :(a) standardising the collection and exchange of HI within and between domains, between MS, including e-health platforms; (b) ensuring data quality, including procedures for internal and external validation of health indicators; (c) undertaking priority setting exercises for HI; (d) addressing ethical and legal issues associated with the collection and use of health data within MS and the EU.

The project was organized in twelve working packages, as follows:

- WP1 Coordination of the project
- WP2 Dissemination of the project
- WP3 Evaluation of the project
- WP4 European Core Health Indicators Monitoring (ECHIM)
- WP5 Harmonized population-based HES
- WP6 Impact of environmental chemicals on health
- WP7 Reproductive, maternal, new-born, child and adolescent health
- WP8 Platform for population-based registries
- WP9 Platform for injury surveillance
- WP10 Building a platform for administrative data on healthcare
- WP11 Integration of approaches into a comprehensive EU information system for health and healthcare monitoring and reporting
- WP12 Evaluation of healthcare systems

Moreover, there were seven horizontal activities, as follows:

- HA1 = Transferability of HI and data for policy making
- HA2 = HI inequality within the EU and MS
- HA3 = Information at regional level (European Core Health Indicators [ECHI], health inequalities) and for specific population groups
- HA4 = Standardization methods for the collection and exchange of HI
- HA5 = Data quality methods including internal and external validation of indicators

HA6 = Priority setting methods in HI

HA7 = Ethical and legal issues in HI.

The objective of the WP8 Platform for population-based registries was to gather, harmonise and disseminate procedures/methods and best practices for population-based-registries as a common platform for the provision of community health indicators of disease occurrence in the population, quality of care and outcomes.

Taking advantage of the existing experience from EUROCISS and EUBIROD projects, two tasks have contributed in the realization of the work: **task 1**, focusing on population health, in particular on disease occurrence (mortality and morbidity, incidence, survival and case fatality rates, and prevalence), and **task 2**, focusing on health care performance, in particular on quality of care and outcomes.

This report is the result of the task-1 expert network and provides a simple tool to support and stimulate the implementation of population-based registries aimed at assessing disease occurrence indicators.

In recent years, thanks to information technology, substantial volumes of data are recorded on hospital admissions and discharges, in-patient care utilization, drug prescriptions, outpatient visits, exemption, general practitioner (GP) databases, surgical operations and invasive procedures. These data, if properly linked, may identify events, which, if checked for quality and validated, can be important sources of information for implementing population-based registries.

“A population-based registry is defined as an organized system that uses observational study methods to collect standardized data to evaluate outcomes for a population defined by a particular disease, condition, or exposure that serve for one or more predetermined scientific, clinical and health policy purposes” [5].

Registries are classified according to how their populations are defined. For example, product registries include patients who have been exposed to biopharmaceutical products or medical devices. Health services registries consist of patients who have had a common procedure, clinical encounter, or hospitalization. Disease or condition registries are defined by patients having the same diagnosis, such as cystic fibrosis or heart failure.

Although registries can serve many purposes, this report focuses on population-based registries created for one or more of the following purposes: to describe the natural history of disease; to assess and monitor the occurrence of the disease; to understand the differences and changes in the natural diseases dynamics between genders, age groups, social classes and ethnic groups; to identify vulnerable

groups; to assess relations between disease incidence, case fatality and mortality; to monitor the consequence of the disease in the community; to trace the utilization and impact of new diagnostic tools and treatments.

This is crucial in order to plan research purposes, in particular on disease causes; to identify gaps in surveillance, develop health strategies and prevention policies, plan future needs for health services and health expenditures, improve appropriate allocation of resources, generate evidence on efficacy and effectiveness of diagnostic and therapeutic interventions.

Recently, the WHO published the Global Action Plan for the prevention and control of NCDs 2013-2020, establishing and strengthening national surveillance and monitoring, including improved data collection on risk factors, morbidity and mortality. The aim of the Global Action Plan for NCDs 2013-2020 is to reduce the number of premature deaths from NCDs by 25% by 2025 through nine global targets, such as: a 25% reduction of premature mortality from cardiovascular diseases (CVDs), cancer, diabetes, and chronic respiratory diseases; a 10% reduction in the use of alcohol; a 10% reduction in the prevalence of insufficient physical activity; a 30% reduction in mean population intake of salt/sodium; a 30% reduction in the prevalence of current tobacco use in persons aged ≥ 15 years; a 25% reduction or containment in the prevalence of raised blood pressure, halting the rise in diabetes and obesity [6].

1.1 The burden of non-communicable diseases

According to WHO report [7], NCDs are the leading cause of premature death and disability globally, accounting for 60% of all deaths worldwide and over 40% of the global burden in terms of loss of healthy life years.

Europe has the highest burden of NCDs globally: 86% of all deaths are caused by diabetes, CVDs, cancer, chronic respiratory diseases, and mental disorders. The burden is increasing further due to a combination of various demographic, epidemiological, economic, environmental, and behavioural trends affecting population lifestyles and associated risks. For instance, demographic changes are a concern because NCDs affect more than 80% of people older than 65 years in Europe, and, according to Eurostat (Statistical Office of the European Community), this age group is projected to almost double in the EU by the year 2060 [8].

For many years, NCDs mortality has been decreasing in the majority of Western European countries and during recent years mortality has also been decreasing in Eastern Europe [9]. However, the absolute number of patients in need of using health services for NCDs conditions does not decrease to the same extent because

prevalence tends to increase, due to an increasing survival and proportion of older people in the population.

NCDs have major economic consequences, as well as human costs: in the WHO European Region, at least 80% of the disability-adjusted life years (DALYs) were due to NCDs [10], including CVDs (21%), cancers (17%) and mental disorders (12%) [11].

NCDs clinically manifest in middle life and older age, after many years of exposure to unhealthy lifestyles (smoking habit, unhealthy diet, physical inactivity) and risk factors (total and low-density lipoprotein cholesterol, blood pressure, diabetes). Even though the clinical onset is mainly acute, NCDs often evolve gradually. NCDs cause substantial loss of quality of life, disability, and life-long dependence on health services and medications.

According to the OECD, it does not appear inevitable that longer life leads to higher healthcare costs. This is one of the reasons why health systems should be largely oriented towards work on preventive actions [12]. Epidemiological studies have shown that NCDs are preventable to a large extent. Different preventive strategies can be implemented to reduce or delay the occurrence and impact of NCDs, increasing the prevalence of favourable risk profile (low risk) of the general population [13], identifying individuals at high risk, and intensifying treatment in those people who have already experienced an event.

In Europe (EU-28 countries), in 2013, mortality for circulatory system diseases was the most common, accounting in 2013 for 34% and 40% of all deaths in men and women, respectively. Diseases of the circulatory system are related to high blood pressure, cholesterol, diabetes and smoking. In people affected by these diseases, the most common causes of death are ischemic heart disease (IHD) and cerebrovascular disease. IHD was responsible for 644,000 deaths, accounting for around 13% of all deaths (99 in women and 176 in men age-standardized rates per 100,000 inhabitants). Stroke was responsible for 433,000 deaths, accounting for 9% of all deaths, (82 in women and 96 in men age-standardized rates per 100,000 inhabitants) [1].

Cancer was the second leading cause of mortality, after CVDs, accounting for 26% of all deaths; more than 1,300,000 people died of cancer. Lung cancer is the most common cause of death from cancer among men; breast cancer is the leading cause of death among women. The percentage of death for cancer was higher in men (30%) than in women (24%) [1].

The third cause of death after CVDs and cancer was represented by respiratory diseases, accounting for 8% of all deaths; more than 400,000 people died from respiratory diseases in 2013, mainly from chronic obstructive pulmonary diseases and pneumonia, but also from asthma, influenza and other diseases. The majority

of deaths for respiratory diseases are age-related, and occur in people aged ≥ 65 years [1].

Coding changes in the international disease classification have posed new challenges for the comparability of disease indicators and produced spurious trends in disease frequency, severity, prognosis and subsequent variations in medical practice, if not properly controlled with the adoption of updated and valid epidemiological methods.

The magnitude of NCDs contrasts with the paucity, weak quality and comparability of data available on the incidence and prevalence of NCDs beyond mortality.

2. OBJECTIVES

The purpose of this report is to gather procedures and methods of different population-based registries, describe opportunities and weakness, provide a general guide and updated methods for the implementation of a population-based registry.

Taking into account advances in diagnostic criteria, treatment and information technologies, as well as experiences and history of different population-based registries, this report provides, as a manual, a standardised and simple model for the implementation of a population-based registry. It recommends to start from a minimum data set (MDS) and to follow a step-wise procedure based on standardised data collection, appropriate record linkage, and validation methods.

Although in many countries data extracted from some sources of information (mortality and HDRs, drug dispensing registries, GP registries, etc.) are now available thanks to the continuing process of computerisation, data are rarely interconnected in order to assess number of events or reliable and comparable statistics. These data, if cleaned, validated through independent epidemiological teams and processed with epidemiological methods, can be used as sources of information for population-based registries.

This report represents a tool to build population-based registries which aim to provide information on disease occurrence, incidence and survival rates and to study the effects of various aspects of services for prevention, treatment and care. Core indicators (incidence, survival rate and case fatality) recommended by the ECHIM Project included in the short list of health indicators, could be assessed [14].

This work is intended for investigators, health professionals, policy makers and data collection staff interested in NCDs or other conditions that require a population-based registration system.

3. POPULATION-BASED REGISTRIES

3.1 *Definition and characteristics*

A population-based registry is an organized system that uses observational study methods to collect all new cases of a disease in a defined population (most frequently a geographical area); data serve for one or more predetermined scientific, clinical and health policy purposes [adapted by 5].

The “**core**” activity is to provide information on **incidence and survival**; a different number of variables can be collected, allowing studying the effects of various aspects of prevention, treatments and caring services.

For some NCDs, population-based registries are the best data source for incidence and survival rates, in particular for those diseases that have an acute onset, such as coronary and stroke events, and injuries. Registries consider both fatal and non-fatal events occurring in-hospital and out-of-hospital, all new cases and recurrent events, in a defined general population, whether treated at home or in hospital, in whichever season of the year or time of the day they may occur, and would also include sudden fatal cases unable to reach the medical service, thus providing estimates of key indicators such as incidence and case fatality rates and survival rates. For other NCDs such as cancer and type 1 diabetes, the definition of onset is an arbitrary concept since it is a continuum of the disease’s natural history; in that case, incidence (new case of diseases) corresponds to the time of the clinical diagnosis, after a patient has presented to medical attention.

The burden of disease and its probable future evolution can be evaluated in terms of incidence and mortality, but other dimensions can be considered, such as prevalence, person years of life lost, quality or disability-adjusted life years: a deep knowledge of the history of a disease may help to project trends into the future and to assess probable effects in changing risk factors.

Focusing on general population, population-based registries may provide a **comprehensive picture of a disease in the community**, highlight problem areas and suggest where treatment facilities are most in need of improvement; they may also provide information systems needed to plan healthcare services, and develop and test which methods are most useful as a basis for preventive actions. This is crucial in order to **plan research purposes**, identifying causes and monitoring progress in prevention, produce annual reports, **orientate preventive actions**, make comparisons among countries in order to achieve better knowledge and more effective interventions, and **support decision making**.

Information from multiple sources contributes to the population-based registry database; therefore it is important to link all the records pertaining to an individual to **avoid duplicate registration**. A **Personal Identification Number (PIN)** is ideal for this purpose; however only Nordic Countries have it. The PIN for each subject is a strong tool in linkage procedures between different sources of data, such as hospital discharge diagnoses, GP records, and death certificates; alternatively, multiple variables (e.g. date and place of birth, sex, residence) may be used for record linkage. This is not possible in the majority of European countries due to privacy and ethical norms and laws, despite the new European law on this matter [15].

Potentially, an individual can have than one cancer or more than one coronary event; with improved survival, this is becoming more frequent. Incidence and survival rates relate to a specific event, so the **new case must be distinguished from the recurrent** (attack rate). The definition of the event should take into account both the ICD codes reported in the hospital discharge diagnoses (main or secondary) or in the causes of death (underlying or secondary) and the duration of the event. In the case of coronary and stroke events, hospital admissions and deaths occurring within 28 days (onset is day 1) are considered to reflect the same event [16, 17] (see section 5.1.8 - B for definition of events).

The quality of the registry data is evaluated by its completeness, validity and timeliness. Completeness of case ascertainment should be as close as 100%; validity, (the accuracy of recorded data) can be increased by checks on recorded data [18]. The rate accuracy is related to the completeness and quality control of data collected for numerator (all events from death certificates, hospital discharge registry, GPs,...) and denominator (census or population under surveillance). Completeness also depends on tracing pauci or asymptomatic subjects treated **outside hospitals** (nursing homes, clinics, GPs). A valid population-based registry should also collect non-fatal events in the target population occurring **outside the area** of surveillance.

Identification of events can be obtained by “hot pursuit” or “cold pursuit”. **Hot pursuit** means identifying case admissions to hospital usually within one or two days of event onset and acquiring relevant information by visiting the ward or interviewing the patient. Information bias is minimised by the hot pursuit approach as information is collected immediately after the event. The process is very expensive for the numerous suspected events collected for validation.

Cold pursuit implies the use of routine and delayed procedures, by means of hospital discharge, and review of clinical and death records. The process is easier and less expensive than hot pursuit; the number of cases studied is typically smaller because discharge diagnoses are more precise and specific than those on admission, but there is a possibility of missing important information when they

are not recorded in the hospital discharge registry. Both methods, hot and cold pursuit, are used to identify suspected events, which are subsequently validated using standardised methods to define diagnostic criteria.

Population-based registry **must be validated**. Validation provides the means to take into account bias from diagnostic practices and changes in coding systems; it traces the impact of new diagnostic tools and re-definition of events; ensures data comparability within the registry (i.e. different sub-populations, different time points, etc.); ensures data comparability with other registries within and between countries. The strength of population-based registries lies in the possibility of **validating each single event according to standardised diagnostic criteria** and collecting disease-specific clinical and paraclinical data [16,17]. This process implies a **great effort in training personnel**, in implementing quality controls for reading and collecting information, for the classification of events according to standardised diagnostic criteria, and for local site visits to assure that standard level are respected and maintained.

Incidence and survival rates from population-based registries have been criticized for the validity of information they provide, in particular when rates are compared in different populations or in different periods of time; a problem common to all comparative studies is the effect of changes in disease classification and coding over time.

Survival, in terms of average number of years lived after a disease, is easy to measure, but it is a “crude indicator” of outcome; years of life are of little value if they are accompanied by disability. **The measurement of health-related quality of life should be part of population based information.**

In some countries, population-based registries have a legal basis (cancer registries); other registries operate on a voluntary basis (coronary and cerebrovascular events), covering populations not representative of the entire country. This means that registry associations have great importance in recommending **common definitions, coding, quality control methods and team training**. A limited geographic coverage is adequate for many descriptive activities, although national data are important to avoid losing migrating subjects or events treated out of the surveillance area. The advantage of using population-based data is that **they relate to the whole community rather than to a single institution or self-selected and atypical subgroups of patients** (those reaching or recovered in a specific centre).

Due to their accuracy in identifying and validating events, results from population-based registries are available usually with a **delay of 3-5 years** in comparison to current administrative data and statistics. All these issues make population-based registries **very expensive**, therefore **this kind of registries can be usually**

maintained only for a limited period, in a defined population of a reasonable size, to answer the specific questions for which they were instituted; local or regional registries may not be representative of the whole country; these are the major limits for the implementation of a population-based registry.

With population-based databases, **descriptive studies** may be conducted examining the differences in incidence and lifestyles; usually these descriptive studies are important to generate hypothesis on risk factors; for example, the role of salt intake in gastric cancer was suggested looking the difference in different countries, studying migrant and time trends [19, 20].

HES and health information surveys can further supplement the information collected from population-based registries with additional details on socio-demographic characteristics, lifestyles, risk factors, and physical/biological measurements.

A population-based registry is intended for health professionals, researchers and policy makers.

3.2 Historical background of population-based registries

An exhaustive analysis of all available registries would have gone beyond the scope of this Work Package, therefore we focused on those registries for high-impact diseases or for specific conditions that could provide any indication on the methodology applied to favour sustainability and implementation of a population-based registry.

For the purposes of this report, we relied on the manuals of operations for population-based registries published or retrieved from the web, on historical research regarding objectives, procedures and methods applied to population-based registries, duration and sustainability of registries, as well as on the practical experience of those involved in registry management. Additionally, the user guide for Registries for Evaluating Patient Outcomes of the Agency for Healthcare Research and Quality [5] and the final report of the PARENT-Patient Registries Initiative were considered [21].

Epidemiologists, public health professionals and health policy makers contributed to this report, with the aim to achieve full coverage of events, completeness of information and validation of diagnoses, so as to provide quality of data assurance and minimize the time needed to obtain the information required.

3.2.1 Cardiovascular diseases

The first experience of a population-based registry in the field of CVDs was the *WHO Myocardial Infarction Community Registers* in 1967 [22]; it was implemented by a group of experts convened by the WHO Regional office for Europe to (a) evaluate the extent of acute myocardial infarction (AMI) in the community; (b) monitor the effect of changes in the management of AMI and different types of intervention; (c) provide an assessment of the validity of mortality statistics; (d) select a pool of patients who could be studied in detail and focus attention on specific problem areas. In order to obtain a statistically sufficient number of notifications, it was decided that the duration of the registration should cover one year, with a 12 months follow-up after the acute attack; each community had to be well defined demographically as census data were indispensable for establishing incidence. All persons in whom there was “any suspicion” that AMI [on the basis of history, electrocardiogram (ECG), enzyme and post mortem results] might have occurred in population who were ≤ 65 years old at the onset of the acute attack and who were resident in the registration area were admitted to the registry. The cut-off point of 65 years was chosen in order to keep the registry to a size that was easy to handle and to exclude older patients with multiple pathologies. The registry examined the incidence of AMI and the influence of smoking, obesity and hypertension on AMI to show which people in the community were specifically at risk. The results of the WHO AMI Registry improved the natural history of ischaemic heart disease: it revealed substantial differences in incidence rates, differences in the time at which death occur in the ability of medical services.

The WHO Myocardial Infarction Community Registers were followed by the *WHO MONICA Project* (MONItoring Trends and Determinants in Cardiovascular Disease) [23], designed to answer key questions arising from the 1978 Bethesda Conference on the Decline in Coronary Heart Disease Mortality: “are reported declines in coronary heart disease mortality genuine? If they are, how much is attributable to improved survival rather than to decline coronary event rates? Are these trends related to changes in risk factors and health care?” [24] It was a very wide project conducted overall the world between mid '80s and mid '90s and allowed for the first time (a) to collect and register - during a 10-year surveillance of 37 populations in 21 countries - 166,000 events in men and women aged 35-64 years; (b) to classify, following the same standardised diagnostic criteria (site and duration of chest pain, evolution of ECG findings, variation of cardiac enzyme values and history of Ischemic Heart Disease -IHD-, and, if performed, necropsy) all suspected events in fatal and non-fatal definite events, possible, ischemic cardiac arrest with successful resuscitation, and insufficient data. An important improvement in the use of standardized diagnostic criteria was the introduction of a quantitative ECG coding system, the Minnesota Code [25]. The main results of the WHO MONICA demonstrated that contributions to changing of IHD mortality

varied, but in populations in which mortality decreased, coronary event rates contributed for two thirds and case fatality for one-third [24]. The extent to these trends was related to changes in known risk factors (systolic blood pressure, total cholesterol, smoking habit and body mass index) daily living habits, health care and major socio-economic features measured at the same time in defined communities in different countries [23].

In the last decade, innovations in diagnostic technologies have facilitated diagnosis at earlier phases in the course of the natural history of disease or in presence of less severe tissue damage. For instance, the use of new biomarkers, such as the routine introduction of new myocyte damage markers (troponins), has required a rethink of the concept of myocardial necrosis and has led to a new and more exhaustive definition of acute coronary syndrome [26-28].

The European Cardiovascular Indicators Surveillance Set (EUROCISS) Project was launched in 2000 by a partnership of EU countries; many of them were collaborating centres of the MONICA Registry and had actively continued the registration of cardiovascular events. The aim of the project was to develop health indicators and recommendations for monitoring the distribution and impact of CVDs in Europe in order to facilitate cross-country comparisons and improve CVDs prevention. Based on the information collected, an updated picture of the existing population-based registries of AMI and stroke in Europe was published, with a detailed description of information sources, data collection, validation methods, including new and old diagnostic criteria [26-28] and indicators assessed in the different population-based registries [29]. Most of these registries, even though deriving from the MONICA experience, basically were not comparable since they collected and validated events with different characteristics (Table 1) [29].

AMI attack rate/incidence, case fatality and prevalence were suggested for inclusion in the ECHIM short list (n.24 and 25) [30]. A second phase of the EUROCISS Project (2004-2007) was launched aiming at (a) developing knowledge, tools and expertise among MS for CVDs surveillance and prevention; (b) preparing the Manual of Operations for the implementation of a population-based registry of AMI/acute coronary syndrome (AMI/ACS) [31]; (c) the implementation of a population-based registry of stroke [32]; and (d) the implementation of CVD surveys for assessment of standardized indicators (prevalence of IHD, heart failure, cerebrovascular accidents and other CVDs, and the identification of a minimum set of questions and exams to be included in the Health Interview Survey (HIS)/HES to evaluate the prevalence of CVDs at European level) [33].

The EUROCISS project, then, provided recommendations on the information sources to be used, on the population size requested to produce stable and reliable indicators, on the age-groups to be considered, on the sample size of events to be validated and on validation methodologies to be implemented.

TABLE 1. REGIONAL POPULATION-BASED AMI/ACS REGISTERS: CASE DEFINITION (FROM EUROCISS FINAL REPORT)

CABG, Coronary Bypass Grafting; ECG, Electrocardiogram; MONICA, MONItoring of trends and determinants in Cardiovascular diseases; PIN, Personal Identification Number; PTCA, Percutaneous Coronary Angioplasty

Country	ICD version	Sources of information		Age range	Linkage mortality / HDR	Validation
		Mortality ICD codes *	HDR ICD codes *			
Belgium Charleroi, Ghent, Bruges	IX, X	410-414, 428, 798, 799	410-414, 428, PTCA, CABG	25 to 69; 25 to 74	name, date of birth	ECG, enzymes, symptoms, MONICA
Northern Denmark	VIII, X	410	410	all	PIN	No validation
Finland	X	410, 411, 428, 798, 799	410, 411, PTCA, CABG	35-85	PIN	MONICA, troponine
France	IX, X	410-414, 428, 798, 799, others	410-414, 428	25 to 64 (until '96) 35 to 74 (from '97)	name, date of birth	MONICA
Germany	X	410-414, 798, 799	410, 411, PTCA, CABG	25 to 74	name, date of birth	MONICA, troponine
Italy	IX	410-414, 798, 799, other	410-414	35 to 74	name, date of birth	MONICA
Norway	X	410	410, PTCA, CABG	all	PIN	no validation
Spain	IX	410-414, 428, 798, 799, other	410-414	25 to 74	name, date of birth	MONICA
Northern Sweden MONICA	X	410, 411	410	35 to 74	PIN	MONICA

CABG, Coronary Bypass Grafting; ECG, Electrocardiogram; MONICA, MONItoring of trends and determinants in Cardiovascular diseases; PIN, Personal Identification Number; PTCA, Percutaneous Coronary Angioplasty

*all codes are presented in the ICD-9 revision to facilitate the comparison

At national level, the Pilot Italian Registry of Coronary and Cerebrovascular event was implemented, covering fatal and non-fatal coronary and cerebrovascular events in the general population aged 35-74 years. It was launched in Italy in 2000, following the MONICA and EUROCISS experiences, coordinated by the Istituto Superiore di Sanità and with the aim of estimating periodically attack rates and case fatality rates of coronary and cerebrovascular events in several geographical areas representative of the country, in order to monitor time trends of CVDs with a major impact in adult population. Current events are assessed through record linkage between two main information sources: death certificates and hospital

discharge diagnosis records; events are identified through the International Classification of Diseases (ICD) codes and duration. Figure 1 shows the ICD codes used in each information source to select fatal and non-fatal coronary events and methods to identify conventional duration ≤ 28 days for each event. Random samples of current events are validated applying the MONICA diagnostic criteria; sample events are classified as definite, possible, and insufficient data based on presence and duration of symptoms, ECG read by Minnesota code, cardiac enzymes, history of IHD, and autopsy; non-fatal events include those classified as definite and possible; while fatal events include those classified as definite, possible and insufficient data [34]; validated events consent to assess the positive predictive value (PPV) for each ICD code of the main cause of death or fatal events and for each ICD code of the first hospital discharge diagnosis of non-fatal events (Figure 2). To calculate the number of estimated events, the number of current events is multiplied by PPV of each specific mortality or discharge ICD code derived from the validation of the random sample of current events (Figure 3). For every 10-years age range, attack rates for both first and recurrent events in the age range 35-74 years are then calculated by dividing the number of estimated events by the resident population; these attack rates are standardized by direct method, using the European Standard Population; the case fatality rate at the 28th day is determined by the ratio between estimated fatal events and total events [34,35]. A similar methodological path is applied to identify fatal and non-fatal stroke events and estimate related attack rates and case fatality in the population (Figure 4) [36].

Figure 1 - Flow-chart summarizing the methodological path to select fatal and non-*fatal coronary* events starting from mortality and hospital discharge diagnosis databases in the Pilot Study of the Italian Registry of Coronary Events

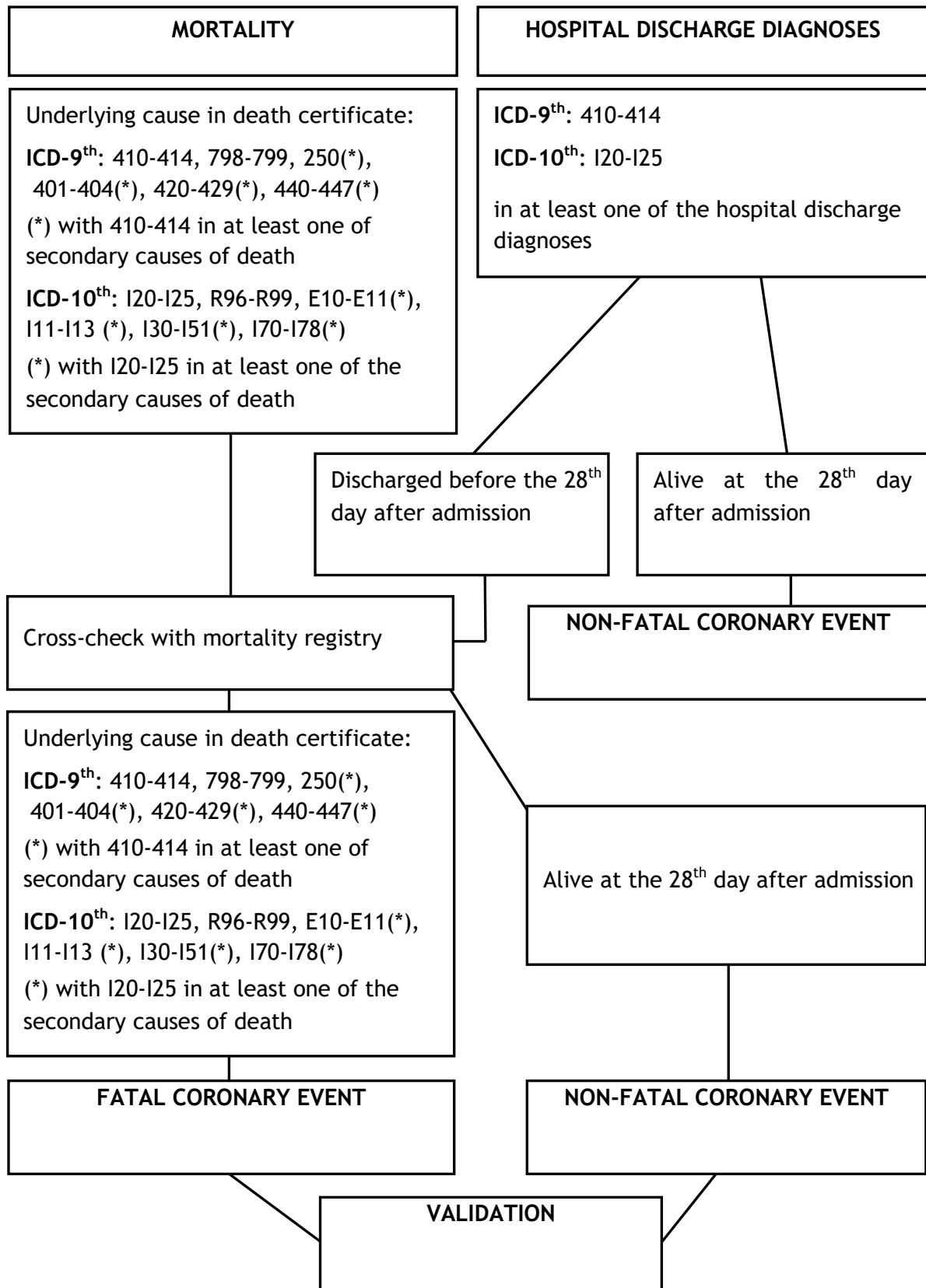


Figure 2 - Positive predictive value for an identified ICD code in the Pilot Study of the Italian Registry of Coronary and Cerebrovascular Events

The *positive predictive value (PPV)* gives the probability of disease for each ICD code

$$PPV = \frac{\text{True (Events)}}{\text{True + False (Events)}} \times 100 \rightarrow \frac{\text{“Validated” Events}}{\text{“Current” Events}}$$

PPV is used to assess the *reliability of each ICD code* in death certificates or hospital discharges

Figure 3 -Number of estimated events for an identified ICD code (fatal and non-fatal events separately)

$$N_{EE} = N_{CE} * \sum (PPV_i * Pr_i)$$

where:

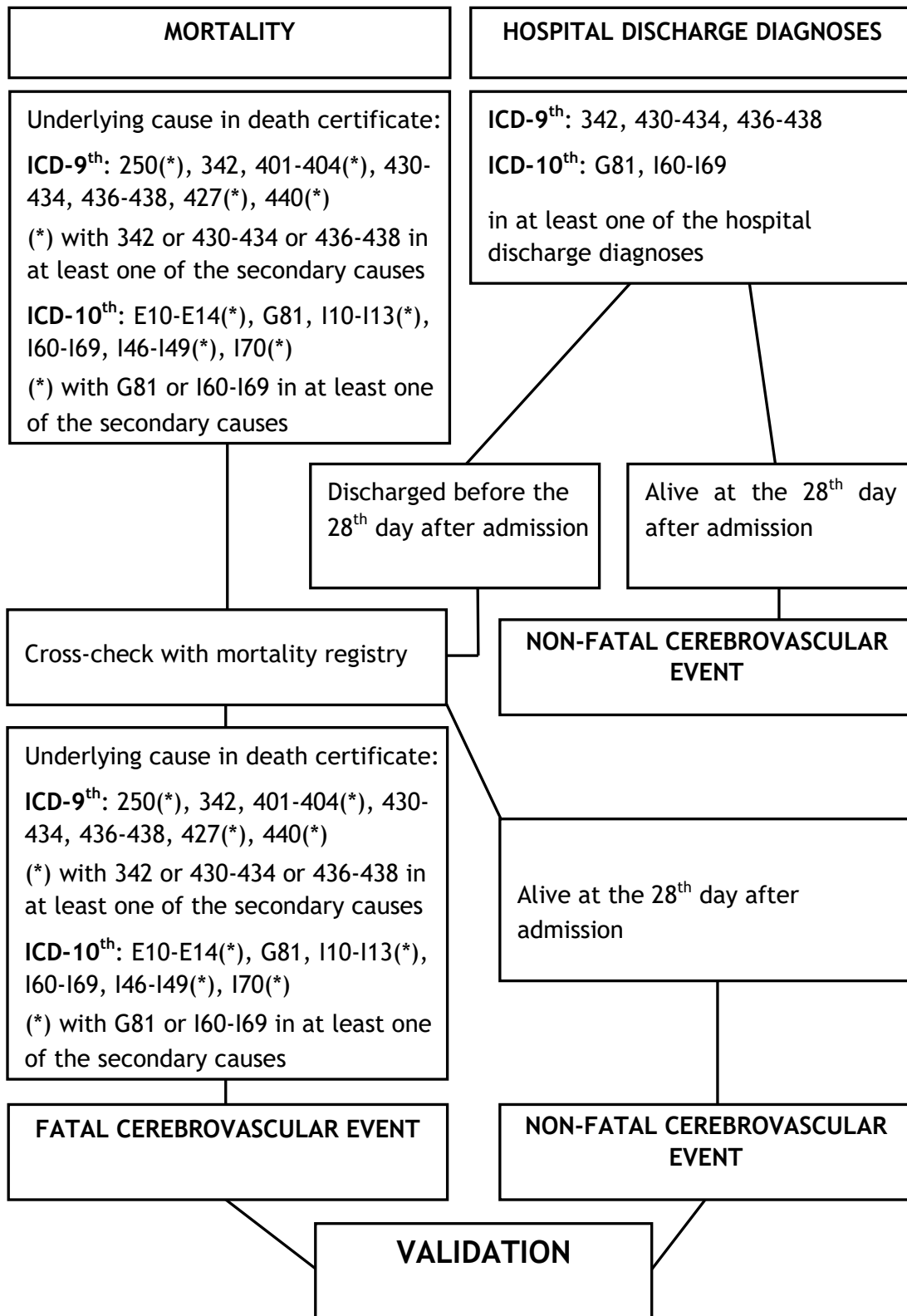
N_{EE} = Number of estimated events

N_{CE} = Number of current events by record linkage

PPV_i = Positive predictive value for the ‘i’-identified ICD code (proportion of events with an identified ICD code validated as positive over the number of total events with the same ICD code)

Pr_i = Prevalence of the ‘i’-identified ICD code

Figure 4 - Flow-chart summarizing the methodological path to select fatal and non-fatal *cerebrovascular events* starting from mortality and hospital discharge diagnosis databases in the Pilot Study of Italian Registry of Cerebrovascular Events



The EuroMed Programme was launched by the EC in 2008 with the purpose of promoting integration and democratic reform across North Africa and Middle East countries neighbours to the EU's South. In the framework of this programme, the Italian Ministry of Health took the lead of initiatives aimed at strengthening health information systems, fighting chronic diseases and developing preventive services. It assigned the project: "A population-based AMI register: assessing the feasibility for a pilot study to implement a surveillance system of acute myocardial infarction (AMI) in Mediterranean countries according to EUROCISS recommendations" to the Istituto Superiore di Sanità. During the Programme, a population-based CVD registry was set up and a validated simplified methodology was developed, able to facilitate the setting up and the implementation of CVD surveillance systems by utilizing a step-wise procedure, as described in the EUROCISS Project [31]. The registry procedure was based on standardized data collection, appropriate record linkage and validation methods, according to scientific criteria defined by MONICA-WHO, the European Society of Cardiology and the American College of Cardiology.

Within the EuroMed Project, an English version was developed and implemented of the software that allows the record linkage of the information sources [mortality data and Hospital Discharge Records (HDRs)] needed for the implementation of the AMI population-based register. Moreover, coronary events, fatal and non-fatal attack rates and case fatality rates were calculated for the population under surveillance. The software is downloadable on request by the website www.cuore.iss.it and is supported by guide-lines for training.

Two countries have been selected for the project: Croatia and Egypt. Training sessions for the AMI registry setting up and implementation were conducted in Zagreb (Croatia) in 2000. In Zagreb, PPVs were not calculated ad hoc on the selected population, but PPVs estimated from the Italian registry were applied to calculate attack rates and case fatality.

3.2.2 Cancer

The history of cancer registration has been a long and slow process, starting unsuccessfully in London in 1728 [37]. The first serious attempt of a census of cancer cases was performed by European countries at the beginning of 1900. Germany started in 1901 by sending a questionnaire to the physicians, and thereafter other countries (Denmark, Hungary, Iceland, Sweden, the Netherlands, Spain, and Portugal) followed this approach. Unfortunately, all these experiences failed because of missing reports and matching data problems. The first real population-based registry was established only in 1927 in Hamburg; it represents the first example of a modern cancer registry, covering a defined population and

using multiple sources of information to identify cases. In the following years, other population-based cancer registries were set up in European countries. In 1950, a WHO subcommittee on the registration of cases of cancer was created and the first methodological guidelines were provided [38].

In 1965, the WHO founded the International Agency for Research on Cancer (IARC), to establish standards for cancer registration, training, publication of registry data and hold scientific meetings. The International Association of Cancer Registries was formed with the purpose to develop and standardize the collection methods across the registries to make their data comparable [18,38].

Nowadays, cancer registries are present in various parts of the world. They cover 5% of the worldwide population, the majority of which in developed countries.

Cancer registries should accept standard procedures and methods, and be part of international associations, such as IARC or ENCR (European Network of Cancer Registries, 1989), and promote the exchange of information between the different registries to improve the quality of data and the comparability between registries.[38]

Registry associations have grown up to deal with issues such as cross notification (cancer patients resident in one area but treated in another) common definitions and coding, quality control procedures and staff training. There are many national associations: UKACR (United Kingdom Association of Cancer Registry), FRANCIM (France Cancer Incidence and Mortality), and AIRTUM (Italian Network of Cancer Registries). Most of them organise training courses, sponsor joint research projects, and hold scientific workshops and meetings. [18]

Data sources of cancer registries are: mortality, hospital discharge diagnosis, treatment facilities (cancer centres, hospitals, clinics), GPs, diagnostic services (pathology departments, clinical laboratories, imaging departments). Each cancer registry collects basic items referred to the person (personal identification, demographic data) and the cancer (morphology, topography, incidence date, diagnosis, behaviour, and grade) pointing on quality of information. The information is collected from these sources by either “active collection” or “passive reporting”: active collection involves registry personnel actually visiting the different sources and abstracting the data on special forms; passive reporting involves healthcare workers completing notification forms developed by the registry or sending copies of discharge abstracts to the registry. An International Classification of Diseases for Oncology (ICD-O) provides the coding for the topography (site) and the morphology (histology) of cancer [39].

Since their establishment, cancer registries have increased their role and, at present, they are considered a useful tool for planning and monitoring strategies and for identifying public health priorities.

The publication of international survival rates by the European cancer registries (EUROCARE Group) showed substantial international variation and influenced policy making in some countries, focusing attention on the reason for the observed differences in survival [40]; survival in terms of average number of years lived after diagnosis is a crude indicator, relatively easy to measure by using the years of life.

3.2.3 Rare diseases

There are more than 6000 known rare diseases (RD); although each RD affects a limited proportion of individuals, it is estimated that, in Europe alone, 29 million people suffer from them [41].

There are two fundamental problems associated with RD: 1) the definition of RD, given that several definitions are in use worldwide and that exiting needs have to be revised as new forms are discovered; 2) the traceability of patients with RD in public health information systems, due to the limitations of the ICD in the identification of all the pathologies. To overcome coding difficulties, an inventory was created, coordinated by Orphanet, the portal for rare diseases and orphan drugs, a consortium of 40 countries of the World [42].

After being ignored for a long period, at present RD are considered as a priority in research programmes and in health policy implementation. Due to RD peculiarity (difficult to seek cases occurring in a well-defined population), only hospital-based registries are concerned with the recording of information on patients examined in specialized hospitals. These databases represent important tools that allow to pool data and achieve a sufficient sample size for clinical research studies, healthcare planning, improvement of care, and quality of life. RD registries have been initiated by many organizations, such as patients and their families, patient advocacy groups, clinicians, national health systems, and biopharmaceutical product manufacturers. According to the data in the Orphanet database (2014), there are 641 disease registries in Europe (40 European, 74 global, 446 national, 77 regional, 4 undefined).

The first initiative in the field of RD was taken by Northern European countries. Sweden established the first centres of expertise for RD in 1990 and a RD database and information centre in 1999; Denmark established an information centre in 1990 and then centres of expertise for RD in 2001; in Italy, a decree on RD came into force in 2001 with the institution of the National Registry of RD patients; in France, Orphanet was established in 1997 with the support of the French Ministry of Health as the portal for information on RD and orphan medicinal products, followed by the first national plan/strategy for RD in Europe (2004). The importance of registries, databases and information networks for RD has been

recognised also at European level in the document “EU Council Recommendation of 8 June 2009 on an action in the field of rare diseases”.

By the end of 2013, 16 countries had adopted a plan or strategy for these RD, and several technical projects on RD have been carried out based on these recommendations, such as EPIRARE (European Platform for Rare Disease Registries), RD-CONNECT (an integrated platform connecting databases, registries, bio banks and clinical bioinformatics for RD research), IRDiRC (International Rare Diseases Research Consortium).

The use of RD registries in Europe is crucial to assess the health status and health outcomes of patients and to monitor the efficacy of health policies by producing health indicators not applicable to all RD, but at least applicable to a significant proportion of them. The choice is to adopt a population-based approach, using indicators such as prevalence, incidence, mortality and fatality rates, and years of life lost [43].

3.2.4 Injury

The first computerized trauma database was established in 1969 at the Cook County Hospital, in Chicago. This registry became the prototype for the Illinois Trauma Registry, which in 1971 began to collect data from 50 trauma centre hospitals across the state. Since then, numerous hospitals, regions, states, and countries have developed trauma registries.

Early trauma registries were characterized by lack of uniformity and dedicated to the epidemiology of traumatic conditions. From 1982 to 1989, a multicentre study, the Major Trauma Outcome Study (MTOS), was coordinated by the American College of Surgeons and collected data from 139 hospitals and over 80,000 injury patients. In 1994, after the success of the MTOS, the National Trauma Data Bank was established. Population-based trauma registries are present in different developed countries (e.g. Canada, Australia, New Zealand, and Israel).

The WHO highlighted the need for injury surveillance. In 2001, the WHO guidelines on injury surveillance were published [44], indicating injury surveillance systems as indispensable to develop effective prevention strategies. In particular, countries need to know about the numbers and types of injuries that occur and about the circumstances in which those injuries occur. The guidelines propose eight minimum dataset variables: identifier, age, sex, intent, place of occurrence, activity, mechanism of injury and nature of injury. The guidelines also recommend a narrative description of the accident that describes how it happened and what the victim was doing.

On the basis of the initiatives for injury surveillance (e.g. in: USA, Australia, Canada, Scandinavia and UK) in 2004 the WHO developed the International Classification of External Causes of Injury (ICECI) [45].

In Europe, to face the problem of injuries, the EC launched the Injury Prevention Programme, which started in 1999 and ended in 2003, when the Public Health Programme (PHP) came into force. Under this programme, in 1999 the European Home and Leisure Accident Surveillance System (EHLASS) was set up by DG SANCO. Within the framework of the PHP 2003-2008, the European Injury Database (IDB) was launched in 13 participating countries. Since 2010, in enforcement of the Council *Recommendation of 31 May 2007* on the prevention of injury and the promotion of safety, the IDB was extended to 22 countries (within PHP 2008-2013) with the goal of reporting on external causes of injuries due to accidents and violence, and become an integral part of the existing exchange programme of Community Statistics on Public Health.

In Europe, injuries represent the fourth most common cause of death, with 235,000 fatalities/year, accounting as the first cause of death and chronic disability in young people. The European IDB is a systematic injury surveillance system that collects accident and injury data from emergency departments (EDs) of selected hospitals. Current mortality data, hospital discharge registry data and other data sources specific to injury areas are used, including data on road accidents and accidents at work. Figure 5 shows the methods used for data collection. Currently, a selection of about 100 hospitals across Europe provides 300,000 cases/year to the database, on unintentional (at home, workplaces, road, sports, etc.) and intentional (violence, self-harm) accidents [46].

With regard to ED surveillance, two different levels of coding are provided: Full Data Set (FDS) and Minimum Data Set (MDS). Both coding systems are compliant with WHO guidelines for injury surveillance. The FDS surveillance has a more analytical coding of the circumstances of the accident and includes a detailed coding of objects or substances involved in the accident and a narrative description of those circumstances; injury data are recorded in the ED of a sample of hospitals. Sampling of cases within hospitals is admitted. It must be random or it may regard only certain injury areas (i.e. only home and leisure injuries), if it is not possible to record all the injuries observed at the ED. The recommended hospital sampling parameters are: minimum 3 hospitals and 8,000 observed cases for countries with a population of less than 3 million people; minimum 5 hospitals and 10,000 observed cases for countries with a population between 3 and 12 million people; minimum 7 hospitals and 12,000 observed cases for countries with a population between 12 and 40 million people; minimum 9 hospitals and 14,000 observed cases for countries with a population exceeding 40 million people. Samples are required to include large and middle-size hospitals, located in urban

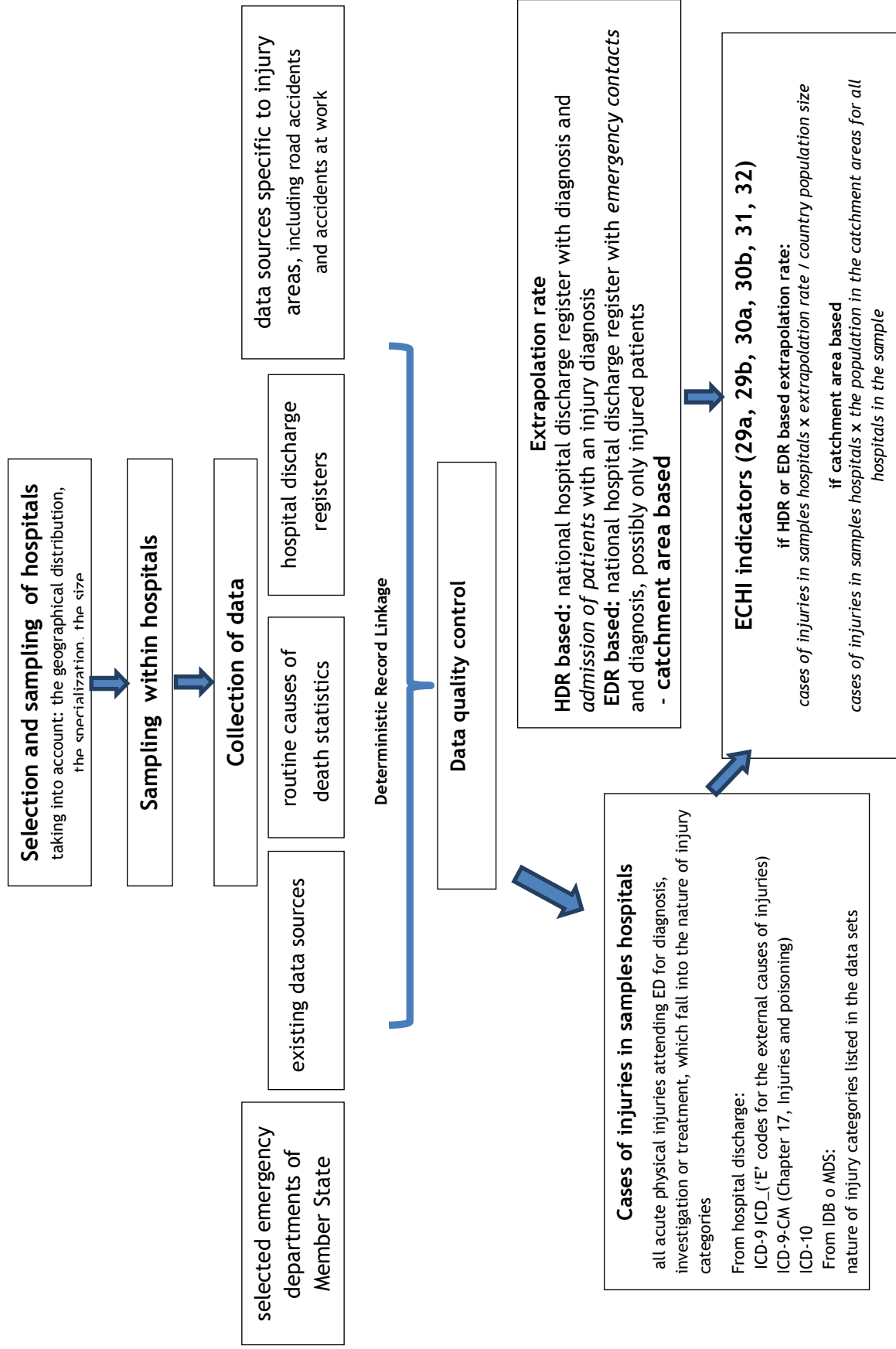
and rural areas and should include hospitals that are accessible by all age groups, including children; if specialized hospitals are included in the sample (i.e. children hospitals or trauma centres) those should be balanced by the inclusion of an adequate number of general hospitals.

The MDS surveillance adopts a simplified coding and does not include objects and substances involved in the accident, nor does it provide a narrative description of the circumstances of the accident. It should include all type of accidents and all the injuries observed in the hospital ED, ideally for the entire country. The MDS surveillance should be representative and include a large random sample taken from the hospital population in the country, at least equal to 10 percent of the total population. In order to be representative, the hospital sampling, even when it is not random, should pay special attention to the geographical distribution of the sample hospitals and to their specialization: the hospital sample should be stratified according to its size; all specializations should be included and the major Regions of the country should be represented.

The IDB is a population-based registry and can complement existing registries with information on the external causes of injuries. It uses data from national health-care registries such as the HDR or ED registers to extrapolate data and produce estimates of national incidence. Deterministic data linkage is provided with HDR records within the hospital sample, in particular to confirm the injury diagnoses recorded at the EDs. Where possible, the linkage is also performed with mortality register for causes of death and with specific data sources related to the information system of the hospital, i.e. labour injuries for social insurance, or road traffic injuries recorded by police forces.

The European IDB data source has been judged as credible and sustainable enough to be included into the health information system and in the ECHI list (2011) [14]. With respect to injuries, there are a few indicators related to home and leisure time injuries - reported by survey or from hospital discharge (indicators 29a and 29b) - and indicators related to road traffic injuries (indicators 30a and 30b), work-related injuries (indicator 31), and suicide attempts (indicator 32). The 29b home and leisure injury indicator covers accidents that have occurred in and around home, in leisure time and at school and resulting in an injury that required treatment in a hospital.

Figure 5. The European Injury Data Base flow-chart



3.2.5 Type 1 diabetes mellitus

The epidemiology of childhood-onset type 1 insulin-dependent diabetes in Europe was studied by the EURODIAB Collaborative Group, established in 1988 as a prospective geographically-defined registry of new cases diagnosed to children under the age of 15 [47]. Twenty population-based registries in 17 countries defined type 1 diabetes on the basis of a clinical diagnosis of idiopathic diabetes made by a physician; date of onset was taken as the date of first insulin injection; anonymous data were submitted to a central coordinating centre for data processing and analysis. More centres joined the group, and most European countries were represented with 6 years in the study period. Capture-recapture methodology, which assumes the availability of independent primary and secondary sources of ascertainment, was used to estimate the completeness of registration [48].

The WHO began the multinational project for Childhood Diabetes (DIAMOND) in 1990. Since then, standardised incidence on type 1 diabetes has been collected within the WHO DIAMOND Project until the year 1999. 114 populations participated, including children aged 14 years or under. Eligible individuals begun daily insulin injection before 15th birthday and were resident in the area of registration at the time of the first insulin administration. Completeness of registration was confirmed by estimating the degree of ascertainment using the capture-recapture method [49]. The rising incidence of type 1 diabetes globally suggested the need for continuous monitoring of incidence by using standardised methods, in order to plan prevention strategy.

The Registry for Type 1 Diabetes Mellitus in Italy (RIDI) was set up in 1997 aiming at recording all new cases in the age range 0-14 years, coordinating the pre-existing registries involved in EURODIAB for the incidence of type 1 diabetes and promoting the establishment of new local registries in uncovered areas (Figure 6).

The RIDI contains basic information on children aged 0-14 years with newly diagnosed type 1 diabetes. Data collected at diabetes onset include personal identification number, date and place of birth, sex, date of diagnosis (defined as the date of the first insulin injection), and municipality of residence. This information established the “national database” of incidence cases; other information such as data sources, ethnicity, height, weight and pubertal status at diagnosis, date and place of birth of parents, number and gender of siblings, family history of type 1 diabetes (parents and siblings), screening for type 1 diabetes, date of diagnosis and vaccination for the following infectious diseases: measles, mumps, varicella, pertussis, roseola, hepatitis B, scarlet fever, are also gathered.

The diagnosis of type 1 diabetes is based on permanent insulin treatment within 6 months from diagnosis with fasting C-peptide levels ≤ 0.20 nmol/l and presence of islet cell antibodies, or glutamic acid decarboxylase autoantibody test (GAD

antibody test). Cases diagnosed as type 2 diabetes or other specific types are excluded.

The registry uses at least two independent data sources for case ascertainment: hospital discharge records, prescriptions, individual National Health Service cards needed by each patient to obtain syringes and strips free of charge, summer camp rosters for diabetic children, membership lists of patient associations, and records of diabetes centres.

The completeness of ascertainment of each registry has been estimated by using the capture-recapture method [50].

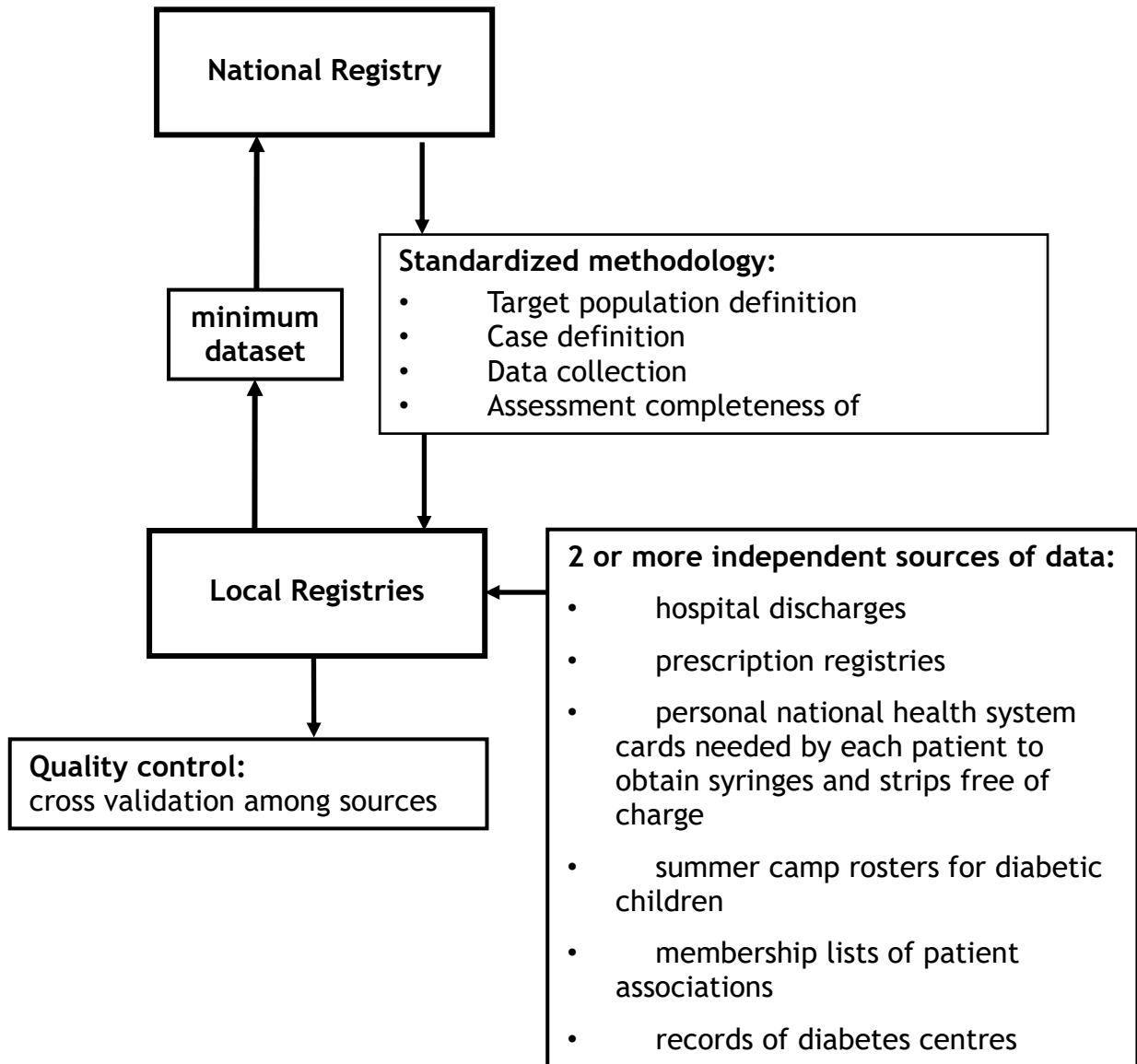
Today RIDI includes a total of 15 local registers: nine regional registries (Valle d'Aosta, Liguria, Marche, Umbria, Lazio, Abruzzo, Campania, Calabria, Sardinia) and six province registers (Trento, Torino, Pavia, Modena, Firenze-Prato, Messina).

Registry ascertainment capabilities range between 91-99%. RIDI's national database includes more than 10,000 cases.

The main results achieved are the standardization of procedures and methods for data collection allowing the assessment of type 1 diabetes incidence, setting up a national database for monitoring and studying the epidemiology of diabetes and comparing data with other areas of the world [51,52]. In ECHIM, there are no indicators of diabetes type 1 in the population aged 0-14 years; it will be possible to assess the "diabetes, register-based prevalence" within the ECHIM indicators of the Work in progress Section.

The European Best Information through Regional Outcomes in Diabetes (EUBIROD) [53] is a European integrated database on type 1 and type 2 diabetes of existing national/regional frameworks, which uses the BIRO technology to automatically and safely generate local statistical reports, in particular on health care performance. The BIRO data set is defined, assessed, and periodically revised by clinical experts, epidemiologists, statisticians, and Information Technology (IT) experts. The EUBIROD survey, conducted across EUBIROD diabetes registers, contributed to assess the consistency of standard definitions with local practices. The BIRO is a system which helps centralize and aggregate databases to a central server, as an essential element for secondary use of health data; for this reason, event definitions, procedures and methods for data collection are not standardised, and each country adopts its own proper clinical judgement and sources of information.

Figure 6 - Flow-chart of Registry for Type 1 Diabetes Mellitus in Italy (RIDI)



3.2.6 Congenital hypothyroidism

The clinical manifestations of congenital hypothyroidism (CH) are often subtle or not present at birth. More specific symptoms do not develop until several months of age. Common clinical features include decreased activity and increased sleep, feeding difficulty and constipation, prolonged jaundice, myxedematous facies, large fontanels (especially posterior), macroglossia, a distended abdomen with umbilical hernia, and hypotonia. Slow linear growth and developmental delay are usually apparent by 4-6 months of age. Without treatment, CH results in severe intellectual deficit and short stature [54].

The Italian National Registry of Infants with Congenital Hypothyroidism (INRICH) is a part of the Italian screening programme for CH. The programme represents an integrated approach to the disease and includes screening tests, diagnosis, treatment, follow-up and surveillance of the disease.

Specifically, the Italian Centres in charge of screening, diagnosis, and follow-up of infants with CH participate in the INRICH, which performs the nationwide surveillance of the disease. The INRICH was established in 1987 as a programme of the Ministry of Health and is coordinated by the Istituto Superiore di Sanità. The aim of the INRICH is: 1) to monitor efficiency and effectiveness of neonatal screening for CH, 2) to provide disease surveillance, and 3) to allow the identification of possible aetiological risk factors for CH.

Information on new cases with CH is collected in the INRICH by means of 3 questionnaires filled in at diagnosis. These include anonymous data concerning CH infants, such as screening and confirmatory laboratory tests, information on demographic data, details on clinical state in the neonatal period, diagnostic investigations (biochemical determinations, radiography of the knee, thyroid scintigraphy, and ultrasound), information regarding pregnancy, birth, and family background, starting and dose of replacement therapy.

The 25 Italian Screening Centres for CH are responsible for contacting the birth clinics and the follow-up centres in order to collect information on new cases of CH and send data to the INRICH. Data is coded and stored in an informed database at the Istituto Superiore di Sanità and results of the Registry are reported in the web site [54].

Over the years, the INRICH has contributed to identify critical points in the screening programme procedures and has therefore contributed to improve diagnosis, treatment and follow-up of affected babies. Moreover, the large amount and the high quality of information collected in the INRICH have provided a unique opportunity for research on this disease [55,56].

3.2.7 Twins

Population-based twin registries are a major resource for genetic and epidemiological studies with a wide range of phenotypes. The largest population-based twin registries are in Scandinavia, but large population-based or volunteer twin registries also exist in Germany, Belgium, United Kingdom and Italy. Evidence of genetic influences on disease, behaviour and other traits can be obtained from twin design in which monozygotic twins, genetically identical, are compared to dizygotic twins. However, findings of significant heritability in one population may not be extrapolated to a different population with different exposure to environmental factors.

The Italian Twin Register (ITR) was established at the Istituto Superiore di Sanità in 2001, when a research project for its implementation was funded by the Ministry of Health. The unique opportunity given by the "fiscal code", an alphanumeric identification with demographic information on any single person residing in Italy, introduced in 1976 by the Ministry of Finance, allowed the creation of a database of all potential Italian twins. To date, this database contains the name, family name, date and place of birth and home address of about 1,300,000 "possible twins". Even though an excess of 40% of pseudo-twins was estimated, this still is the world's largest twin population ever collected.

The database of possible twins is currently used in population-based studies on multiple sclerosis, Alzheimer's disease, celiac disease, and type 1 diabetes. A system is currently being developed for linking the database to data from mortality and cancer registries. In 2001, the Italian Government, through the Ministry of Health, financed a broad national research programme on twin studies, including the establishment of a national twin registry [57].

Since its establishment, this registry focused on a continuous update of the existing information, on the one hand, and on new phenotypes and sample collection, on the other hand.

For twins' enrolment, three different strategies are usually followed. The first, which corresponds to a 'population-based approach', consists of sending, by mail, an enrolment kit to all twins of a specific age group living in a specific area. This kit contains a standardized questionnaire, the informed consent form to be signed by the twin, or the twins' parents in case of minors, and an informative letter on the ITR research activities. The second, which corresponds to a 'hospital-based approach', relies on linking the ITR database to disease registries that could be locally or nationally based, in order to detect potential twins to be enrolled in studies of specific diseases. The third approach consists of voluntary requests for

enrolment by twins who have become familiar with the ITR through the website [57,58]. These volunteers also receive the enrolment kit.

The enrolment questionnaire collects demographic information not available in the original database, such as occupation and education; it includes a set of questions on physical resemblance from which zygosity is derived, and a few other items such as current weight and height. To date, data has been collected through paper questionnaires. The questionnaires are either sent by mail or given directly to the twins if a face-to-face meeting is included in a specific study. The questionnaires are then scanned using a character recognition system and data is stored in a central database and server. Twins are then re-contacted according to inclusion criteria (e.g. zygosity, targeted age groups, geographic area) or target outcomes requested by specific studies (e.g. complex traits, specific pathologies, discordance for trait/pathology, etc.).

Up to now, about 28,000 twins participate in the ITR research activities. In the area of behavioural genetics, most efforts have been directed to psychological well-being assessed with self-reported tools. Research on age-related traits continues with studies on atherosclerosis development, early biomarkers in mild cognitive impairment, and relationship between lifestyles and mutagen sensitivity. A valuable key resource for the ITR is the possibility of linking twin data and disease registries. This approach has yielded several important results, such as studies on the heritability of type 1 diabetes, multiple sclerosis and celiac disease. The ITR bio banking has grown in size and know-how in terms of implementation of both technical issues and ethical procedures. Furthermore, attitudes toward bio-bank based research, together with willingness and motivation for donation, are being investigated [58].

3.2.8 Maternal mortality

Maternal mortality (MM) is a crucial indicator of the health status and appropriateness of health services in a country. Maternal deaths involve young women in good health and have a devastating impact on families, health professionals and communities. Around 50% of maternal deaths could be prevented by systematically analysing their causes and correcting them. The link between information and response has been often used in maternal surveillance to get data other than simple figures, but only a few countries with advanced health systems set up enhanced MM surveillance systems.

At international level, the methodology to assess maternal mortality varies considerably [59]. Mortality Registers have been used for international comparisons of maternal mortality ratios (MMR), but are not sufficiently accurate. International reporting - limited to current vital statistics analysis - fails to detect the overall

magnitude of the phenomenon, and also the correction factors and the algorithms applied by WHO to routine maternal mortality data can be misleading [60,61]. Only incident reporting and confidential death enquiries provide the opportunity to estimate accurate MMR, detect the effectiveness of obstetric care and support the identification of the training needs of health professionals.

In Italy the Ministry of Health has supported a number of multi-regional projects [62] coordinated by the Istituto Superiore di Sanità with the objectives of collecting reliable data on maternal mortality and promoting the continuous education of health professionals involved in maternity care in order to prevent and limit avoidable outcomes resulting from complications of pregnancy, childbirth and postpartum, and establish health priorities.

The Italian Surveillance System adopts a dual approach in assessing maternal deaths:

1. Record linkage between hospital discharge databases and cause of death registry (vital statistics analysis)
2. Notification of maternal deaths and confidential enquiries on incident cases (active surveillance)

1. *Vital statistics analysis methodology*

A retrospective methodology based on vital statistics linkage began in 2008 and adopted the WHO ICD-10 maternal death definition. The study population covers all resident women aged 11-60 years, with one or more hospitalizations for pregnancy or any pregnancy outcome (spontaneous abortion, induced abortion, ectopic pregnancy, stillbirth or live births).

Figure 7 - Vital statistics linkage



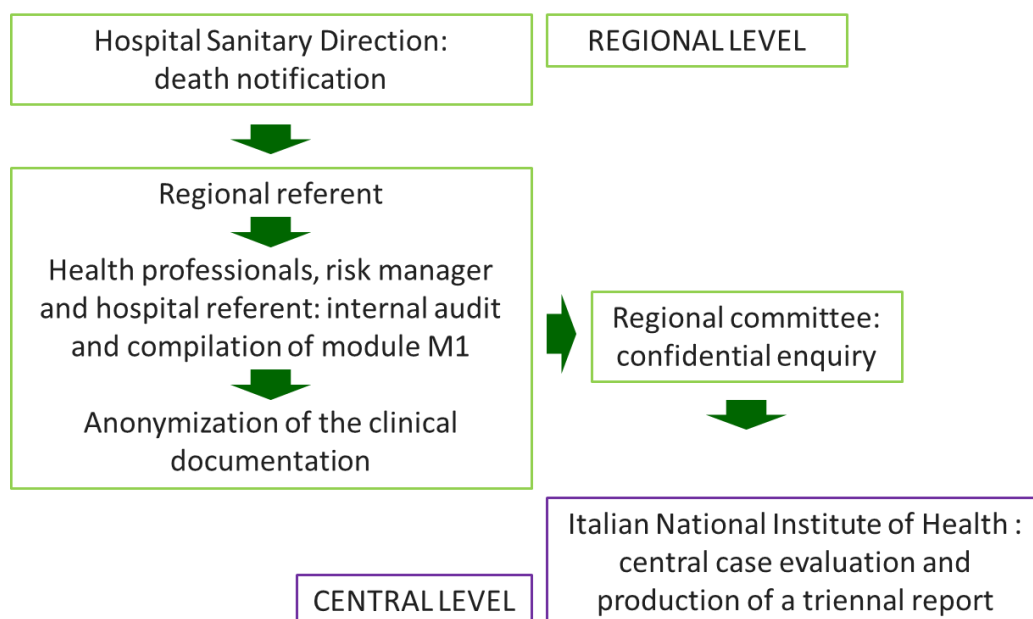
Women were identified by selecting every hospital discharge database reporting one or more protocol diagnosis or procedure code. Selected women were those who had died due to any pregnancy outcome within 365 days from the

outcome, and identified through the record-linkage procedure, or by the death certificate, or by the hospital discharge database. The interval time of 365 days was calculated with reference to the last hospital admission during the pregnancy. The attribution of the underlying cause of death was decided by a clinician at regional level and by a group of experts at the Istituto Superiore di Sanità. Additional procedures were also used, such as the check of multiple diagnoses in the case of more than one hospitalization, and the retrieval and analysis of medical records. Vital statistics analysis allowed computing the number of incident maternal deaths, attributing most of their causes and detecting for the first time a MMR underestimation of 60% in 8 Italian regions covering 75% of total new-borns [63]. The first national MMR estimate will be available in 2017.

2. Active surveillance

Perspective surveillance through incident case reporting and assessment by experts has started in the country since 2012, with the objective of generating the necessary information to outline realistic and practical actions to accelerate progress towards reducing preventable maternal severe morbidity and mortality.

Figure 8 - Flow-chart summarizing the operational aspects of the active surveillance system

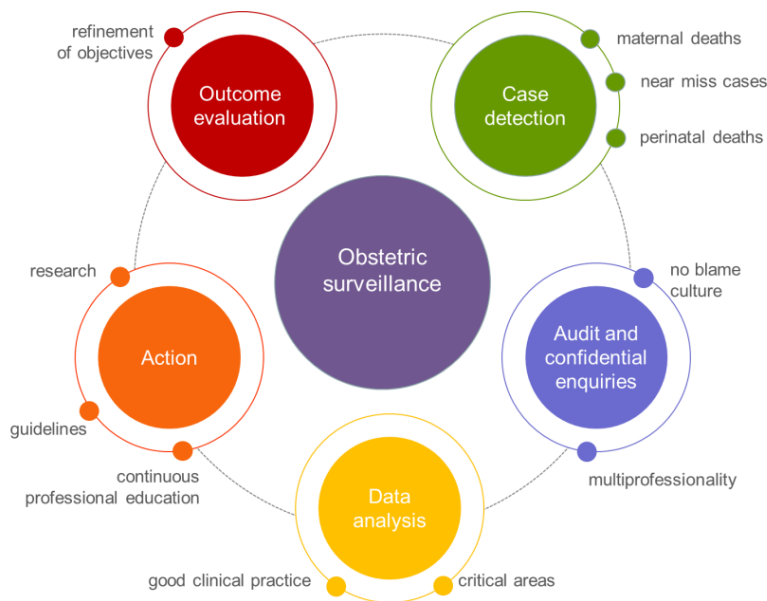


The enhanced surveillance system is based on a population-based approach. Maternal deaths are notified by all public or private health facilities with an obstetric unit and/or an intensive care unit and/or a coronary care unit and/or a stroke unit. A regional coordination unit in each participating region coordinates the selection of a reference person (a motivated medical doctor or midwife) in every facility and formally appoints a multidisciplinary regional expert committee responsible for the confidential enquiries. The committee includes obstetricians, anaesthesiologists, midwives, pathologists, risk managers and epidemiologists of

recognized authority. All hospital reference persons and a representative of the clinical risk management network for each facility received a residential training on the operational aspects of the surveillance system and a training package for cascade teaching to all professionals involved in the women's assistance once back in their health facilities. Every maternal death is notified by the Hospital Sanitary Direction of the facility where the death took place to the regional unit coordinator, who is asked to verify the achievement of an internal audit in the health facility within one month from the death. The audit is facilitated by the risk manager of the facility, in collaboration with the local reference person, and involves all health professionals who assisted the deceased woman. During the audit, a provided anonymous form is filled in to describe the complete clinical history of the patient. The form and the woman's medical records, which are made anonymous, are delivered to the regional expert committee for the confidential enquiry. The results of the enquiry and the complete documentation of the case are then transferred to the Istituto Superiore di Sanità, where a further central review and analysis is carried out to ensure national consistency of classification of maternal deaths. If the results of the central assessments differ from those of the regional committees, the cases are discussed jointly by the national and the regional committees to reach a conclusive shared opinion about the cause of death and the definition of quality of care. The final shared assessment includes the classification of the maternal death, the attribution of its cause and the definition of the quality of care that is described as: appropriate with unavoidable outcome; improvable with unavoidable outcome; substandard with avoidable outcome. The system has been promoting the participation of health professionals and the development of a "no blame" culture through confidentiality of reporting.

This dual approach to investigate and monitor maternal mortality, including vital statistics analysis and a confidential enquiry system, is the best option for case ascertainment and prevention of avoidable maternal deaths. The surveillance enhanced system is based on an audit cycle (Fig.9) that starts with the identification of cases and their systematic data collection, includes a critical analysis to generate recommendations for clinicians and policy makers and implement improvement actions, and closes with an impact assessment. The core of this active surveillance approach is the lessons learned and the actions taken on the results, with the aim of saving more women's lives and at improving quality of maternity service. Up to now, the active surveillance system covers 77% of Italian new-borns and involves almost 400 public and private health facilities.

Figure 9 - The surveillance cycle



3.2.9 Implanted prostheses

Registries for implanted prostheses can be designed for many purposes; they monitor the performance of implanted medical devices, and keep track of procedures, surgical techniques, hospitals, and patient characteristics. The endpoint of an implanted prostheses registry is the revision procedure.

Any medical device placed in Europe must comply with the relevant legislation, in which three types of medical devices are outlined: general medical devices, implantable medical devices, and “in vitro” diagnostic medical devices. Today’s medical devices are becoming ever more sophisticated and innovative. Population in Europe is ageing and it is estimated that in 2060 there will be twice as many Europeans aged 65 or over as we have at present, and this will increase the importance of medical devices for public health and medical care, as for example in joint replacement surgery.

As concerns joint replacement surgery, several implanted prosthesis registries have been set up over the last forty years. The Swedish Hip Arthroplasty Registry (SHAR) started in 1979 and is one of the first examples of joint replacement implants registry’s effectiveness. The SHAR is funded by the government, has a patient coverage of 98% and a hospital coverage of 100% (2009) [64].

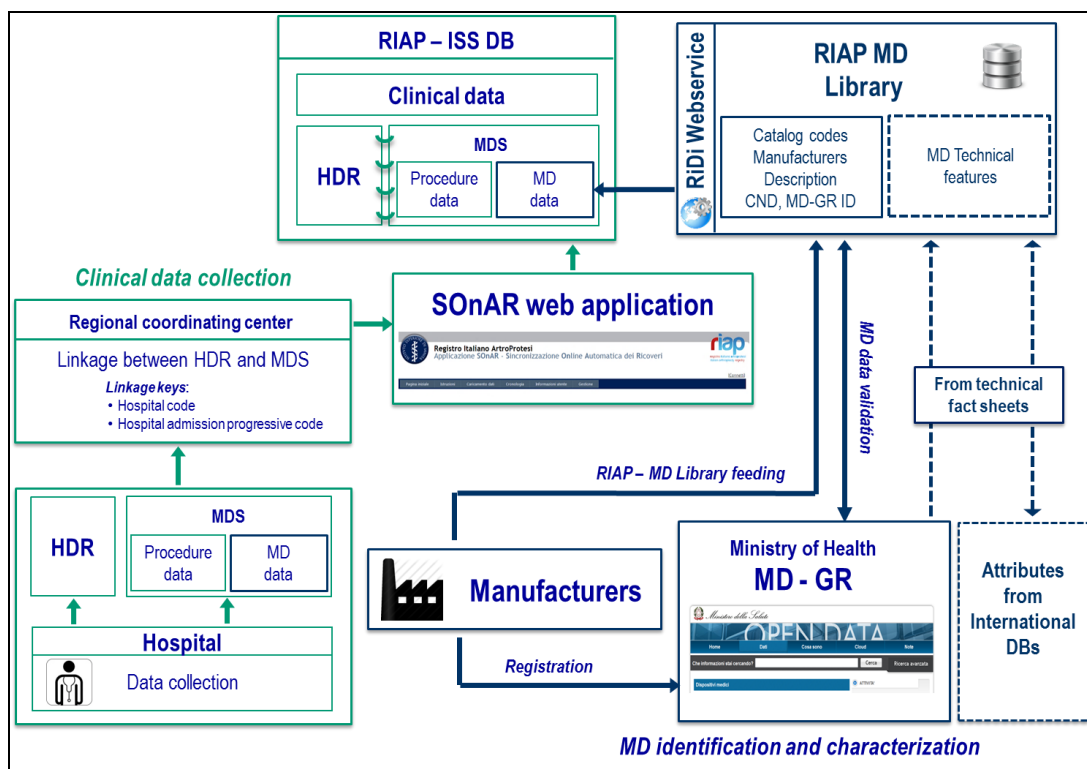
In the same field, the Italian Arthroplasty Registry Project (RIAP) started in 2006. It is based on three pillars: 1) to be a federation of regional registries coordinated by the Istituto Superiore di Sanità; 2) to use HDRs integrated by an additional MDS of information describing procedures (operated side, type of procedure, previous

procedure, diagnosis, surgical approach, fixation method) and implanted medical devices (manufacturer's name, device catalogue code and lot number); 3) to punctually identify every component of the implant using a dedicated library (RIAP Medical Device Library) built in close cooperation with manufacturers, and to describe the device by a list of attributes provided in cooperation with international databases and selected in the General Repository of the Ministry of Health. At present, RIAP collects data from hip, knee, and shoulder procedures; data collection for ankle replacement procedure will start soon.

The activity of RIAP involves the voluntary participation of the various subjects: Italian regions, hospitals, medical devices manufacturers, patients, and scientific societies. RIAP collects and processes only data for those patients who have signed the informed consent. RIAP has been funded by the Italian Ministry of Health, Directorate General of Medical Devices and Pharmaceutical Services, as a tool to support post-marketing surveillance and vigilance activities for medical devices. The endpoint of RIAP is, therefore, the implant revision (removal, partial or total replacement of device), and patients recall in case of failure of the implanted medical device [65].

The RIAP data collection model can be divided into two flows: the clinical data collection (*green*) and the medical device identification and characterization (*blue*) (Figure 10) [66].

Figure 10 - RIAP data collection model



HDR= Hospital Discharge Record; MDS= Minimum dataset; MD= Medical Device; DB= Database; GR= General Repository

Clinical data collection flow (Figure 10, green, left side). It concerns the HDR and MDS (procedure data and MD data) data collection, the linkage between HDR and MDS performed by the Regional coordination centre, and the transmission of linked clinical data to the RIAP - ISS DB by the dedicated web application SOnAR.

MD identification and characterization flow (Figure 10, blue, right side). Particular attention is paid to the punctual identification of every component of the implant: to this aim, the RIAP MD Library has been implemented and continuously fed by manufacturers. RIAP receives the dataset about MD identification data from manufacturers, performs data validation by comparing the received dataset with the Ministry of Health MD - GR (General Repository), and feeds the RIAP MD Library with the validated data. Manufacturers receive a feedback about the results of data validation. Using the RIAP MD Library in data entry procedure decreases transcription errors.

3.2.10 National Italian Transplant Centre

The Italian experience in the organization of transplantation procedures may represent a relevant example of an internal development at national level, combined with a strengthening of international collaborations. These results can be attributed to the creation of the Italian Transplant Centre and to the on-going European process that is leading to the awareness of the importance of a close collaboration between already existing organizations that operate in the transplant field [67].

The National Transplant Centre stems from law n.91 of 1 April 1999 as a technical body of the Ministry of Health. The Centre is located at the Istituto Superiore di Sanità; it was entrusted with the national coordination of the activity of donation, procurement and transplantation of organs, tissues and cells. Its aims are: monitoring of donations, transplants and waiting lists through the information system; definition of guidelines and operational protocols; allocation of organs for urgent cases referred to the national catchment area; definition of the parameters for quality auditing and outcome of transplantation structures; promotion and coordination of the relations with foreign institutions in this sector [68].

3.2.11 PARENT

The PARENT (Patient Registries Initiative) was a Joint Action (JA) aimed at providing MS with guidelines and recommendations for the improvement of patient

registry quality and interoperability readiness, encouraging and improving the use of data for secondary purposes.

The PARENT JA has succeeded in producing a registry of registries (RoR), integrated in the digital European infrastructure.

Over the course of the PARENT JA, two complete versions of the RoR were created. The first was based upon an online questionnaire, which collected information on European patient registries, including information from registry owners and administrators, data sources, uses and needs, methodologies, availability, comparability, and potentiality for secondary use of registry data.

The second version of the RoR expanded upon these functionalities, with the addition of an informative homepage that helped in dissemination activities. The second version also included the ability for registry holders to nominate new organizations and registries as well as edit their data, with links to the Registry Guidelines wiki tool. This version also allows registered users to access a dynamic assessment and comparison tool. The tool is accessible at web page <http://www.parent-ror.eu/#/registries>.

The Methodological Guidelines and Recommendations for Efficient and Rational Governance of Patient Registries enable registry holders, governance bodies, financing institutions, researchers and others to have a comprehensive overview on the necessary activities and structures governing the registries, so to provide high quality data and to re-use data. A wiki tool was created, as a means of offering the guidelines to the users in a searchable and structured way.

In the document “Initiative for Patient Registries - Strategy and Pilot Phase”, the European Medicines Agency (EMA) recognised the PARENT JA RoR and the guidelines as important sources of information for the identification and evaluation of existing data sources.

The PARENT JA outlined clear policy objectives: (1) to strengthen the use, usefulness and suitability of registries for Health Technological Assessment-based (HTA) patient and consumer safety promotion and monitoring; (2) to propose the placement of interoperable patient registries on the national and EU eHealth strategies and roadmaps; (3) to support European Reference Networks and RD services and research, such as National Contact Points for the support of patient registry collaboration [69].

4. SURVEILLANCE

Surveillance is the on-going, systematic collection, analysis, interpretation and dissemination of HI to health professionals and policy makers. Surveillance is

defined as a continuous, and not occasional or intermittent activity; it is used to monitor diseases and health conditions [70].

Surveillance is conducted in many ways, depending on the nature of the event under surveillance, healthcare and information infrastructures, population involved, available resources, and information needs. For many years, mortality statistics have been the main tool for comparing health and disease patterns among countries, and today they still remain the only source of information for a few countries. Since the '50s, causes of death have been registered according to the ICD. Different classification of diseases within versions and different methods of ascertainment have led to comparison problems between different ICD revisions and/or similar versions among countries.

Currently, surveillance can be performed using current databases routinely collected: hospital discharge diagnoses, GP data sets, healthcare data (drug prescription, outpatient visits, exemption). These databases are created to supply payments to health services providers within public or private healthcare delivery systems [71]. Events may be identified with appropriate procedures, using mortality data, HDRs, and other sources of information. These databases are cost-effective, cover the whole country, include all age groups and collect large number of events. The main objective of studies using administrative databases is to produce relevant statistics in order to plan health services and healthcare expenditure, and to provide data on mortality, causes of death and hospital admissions at international level. These data are not primarily planned for research purposes, but they are increasingly used in epidemiological research. Their strength lies in the fact that they cover the whole country, and completeness is close to 100%. On the contrary, their weakness lies in the fact that data are not standardised in the disease specific data collection and that available clinical and lifestyle data are limited. If used in research, routine databases need to be carefully validated [72-75].

Even so, case series from hospital-based registries provide important clinical information. A hospital-based registry collects information about hospital patients through surveillance of admission and discharge records. In particular, it provides detailed information on diagnostic and therapeutic procedures and on risk factor levels at hospital admission. Hence, the primary objective of this type of registers is to assess length of stay, in-hospital treatment, healthcare performance and outcome [76].

HES, HIS and Population-based Registries are the main tools for epidemiological surveillance in the population.

HES is a population-based survey that uses random samples of the country's general population; 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 (e.g. lipids, fasting blood glucose, haemachrome) assayed in a centralised laboratory under the quality assurance system and the supervision or control of an international reference laboratory [usually WHO Reference Centre or Centres for Disease Control and Prevention (CDC)], standardised questionnaire(s) (e.g. positive history of chronic diseases, life styles such as smoking habit, physical activity, diet, pharmacological treatments, family history, physical performance, cognitive function).

HIS is a survey in samples of the general population that includes interviews on health characteristics (perceived health, diseases, disability) health related behaviour (e.g. smoking habit, physical activity), and use of health services. It is based on face-to-face interviews and self-administered questionnaires, telephone interviews, postal surveys.

HIS are used to collect information on self-reported and perceived health status, health determinants and health care; HES provide objective measurements of health related outcomes, but they have high costs and are time consuming. Both HIS and HES, if conducted adopting proper standardized and harmonized approaches, may produce comparable and reliable data.

Many countries conduct HIS, but less countries have the opportunity to conduct HES. The latter are primarily used for monitoring prevalence of disease, prevalence of risk factors (health behaviour, social network, environmental risk factors) and of disease consequences (disability in physical performance or cognitive function, unemployment).

4.1 Data sources for population-based registries

Only few countries have established disease-specific population-based registries that ensure a more precise and valid monitoring of diseases. These registries derive from a variety of current available data from different sources, but the identification of suspected events from these sources of information requires a further level of processing to ensure comparability (validation).

Data sources used in registries depend on the type of disease under surveillance and related-disease events; for example, to monitor NCDs in the general population, at least the following sources of information should be available: mortality records with death certificates; HDRs with clinical records, autopsy registry, nursing home and clinic, emergency and ambulance services, GP clinical records, drug dispensing registry, exemptions.

4.1.1 Death certificates

Death certificates provide complete data on fatal events and are collected in a systematic and continuous way in all countries. Mortality statistics are easily accessible but are usually published in a detailed and complete form after 2-4 years.

The format of death certificates varies across countries: it generally includes personal identification data, date and place of birth, date and place of death (i.e. nursing home, hospital, home), residence, and underlying cause of death (only few countries report underlying, immediate and contributing causes of death). Death is coded according to the International Classification of Disease and causes of death. Problems of temporal and geographic comparisons derive from the different versions of the ICD adopted over time (7th, 8th, 9th, 10th revision) and from different coding practices in each country. Furthermore, diagnostic criteria for coding death certificates are not defined at international level and the ICD versions are updated every 10 years by WHO.

The reliability of mortality data depends on the completeness and accuracy of the vital registration system, as well as on the completeness and accuracy of the registration and coding of causes of death. When the proportion of deaths coded as “unknown cause of death” is higher than 5%, cause-specific mortality data should be used with caution. The accuracy of the recorded causes of death depends on the autopsy rate. This rate varies largely among countries and over time. In some countries, the autopsy rate has declined in recent years, which may be a problem for the use of mortality statistics of some disease.

4.1.2 Hospital discharge and medical records

HDRs give the number of hospitalisations in a definite period of time, which are important for purposes related to management of resources, costs and health services. Hospital discharge data are available in most EU countries; in some countries, data are aggregated without detailed information on age and sex distribution and without separate diagnostic categories (e.g. all cardiovascular diseases, instead of AMI, stroke, heart failure, atrial fibrillation).

HDRs include personal data, admission and discharge dates, type of hospitalisation (urgent, ordinary or transfer from other structures), main procedures (e.g., surgical intervention), discharge diagnoses, condition at discharge (at home, transfer to other structure, dead), and information used for reimbursement (e.g., DRG). Hospital discharge diagnoses are coded by ICD codes (currently ICD-9 or ICD-10). For some countries, only a limited number of diagnoses are coded. Use of HDRs and medical records can generate some difficulties: a) problems in assessing a specific event may arise when an acute event is followed by a period of

rehabilitation in a different hospital or clinic, or a transfer to other wards, as the event could be counted more than once; b) clinical information and medical records reported in hospital documents must be seen and renewed for event validation; c) HDRs are not always validated on a routine basis, and validation studies are necessary to check their diagnostic quality; d) the validity of HDRs may vary according to the geographical region, the type of hospital or clinic and patient characteristics; e) hospital admission policies may vary over time and place, e.g. the registration of the most severe cases, or of the deaths occurred shortly after arrival to hospital may differ among hospitals. HDRs may also include patients not resident in the area under surveillance; f) the adoption of new diagnostic techniques may cause major changes in event rates estimated from HDRs; g) a further problem may derive from the use of Diagnosis Related Group (DRG). In some countries, hospital reimbursement is based on the DRG tariff system, which is built on equal-resources criteria and aggregates events in major diagnostic categories; i) in order to assess the occurrence of events, HDRs from all hospital departments of the geographical area under surveillance should be used.

4.1.3 Autopsy registry

Not all countries routinely perform an autopsy on suspected or sudden deaths. Autopsy is performed on violent deaths or on deaths occurring in hospital when clinical diagnosis is undetermined. The former is performed by forensic medicine specialists, the latter by pathologists of the hospital where the death occurred. Data from autopsy registries refer therefore to a low percentage of deaths, but provide a more valid diagnosis to complement the information reported on the death certificates.

4.1.4 Nursing home and clinics

Nursing home and clinics mainly provide data on cases among elderly patients who sometimes get care from these institutions without being admitted to an hospital. Therefore, information on events occurring in the nursing home may be critical, especially if the registry covers elderly patients up to 84 years of age.

In some countries, rehabilitation after an acute event is provided by the rehabilitation clinic, which may give information on patients who have received acute care outside the region.

4.1.5 Emergency and ambulance services

Data provided by emergency and ambulance services is useful to integrate information for registry implementation, since patients dying from sudden death or experiencing fatal events are not always able to reach the hospital. These services can provide data otherwise not obtainable, such as ECG during the acute phase of the event, blood pressure measurements, level of consciousness and muscular deficit at the time of event occurrence in pauci-symptomatic patients recurring to emergency services. The need for very urgent medical treatments often makes information partial, but the integration of this data with other data from other sources of information contributes to the implementation of the registry and event validation.

4.1.6 General practitioner medical records

GPs provide information on those events that do not reach the hospital and for those patients who are hospitalised outside the area of their usual residence; GPs can provide clinical data and thus integrate information from other sources (HDR, death certificate, etc.). GPs datasets may also provide an adequate coverage for prevalence of NCDs. This network operates in a few countries (e.g., the Netherlands, UK, and Italy).

The GPs network may be affected by selection bias, as usually only volunteer GPs participate in studies. For this reason, data from GP network requires integration with other different sources of information and validation.

4.1.7 Drug dispensing registries

In some countries, patients may receive comprehensive drug reimbursement under their national healthcare system, so drug prescriptions can serve as a proxy for disease.

4.1.8 Other sources

Disease exemption, outpatient visits, laboratory tests, radiological and anatomic-pathological records. The more linked are the sources of information, the lower shall be the probability of missing cases.

5. METHODS

As we have underlined several times in this report, “a population-based registry is an organized system that uses observational study methods to collect uniform data

(clinical and other sources), and to evaluate specified outcomes for the general population that serve predetermined scientific, clinical, or policy purposes” [5]. Studies derived from population-based registries can provide a real-world view of disease occurrence in the population, clinical practice, population or patient outcomes, safety, as well as clinical, comparative, and cost-effectiveness, and may serve a number of evidence developments and decision-making purposes.

To build a population-based registry, some criteria, rules and different steps should be respected.

The following part of this report shall be dedicated to practical design and operational issues, evaluation principles, and good practices.

5.1 Planning a registry

Registries vary in size, scope, and resource requirements. Population-based registries may cover a population large enough to produce stable and robust disease rates. They may target rare or common conditions. They may require the collection of limited amounts of variables, operate for short or long periods of time, and be sustained through financial support. The scope of a registry may be adapted over time to reach broader or different populations. Registries require good planning in order to be successful.

When planning a registry, the following initial steps are desirable: (1) formulating the purpose of the registry; (2) determining if a registry is the appropriate tool to achieve the purpose; (3) identifying key stakeholders interested in the disease registry; and (4) assessing the feasibility of a registry. Once a decision is made to proceed, the next considerations are strictly related to the methods to be adopted, specifically in relation to (5) team building; (6) governance; (7) definition of the scope and rigor needed; (8) definition of data set, outcomes, and target population; (9) development of a protocol; and (10) development of a project plan.

Periodic evaluations of the registry ensure that the objectives are met. This is particularly important for registries that collect data over many years. When objectives are no longer met, or when diagnostic criteria of disease are changed, the registry needs to be adapted, otherwise the collection of new data should be stopped [77,78].

5.1.1 Formulate the purpose of the registry

The first step in planning a registry is the formulation of a clearly defined purpose and rationale; this makes easier to evaluate whether the registry is the right

approach to obtain the data of interest [79,80]. A defined purpose helps to clarify the data needed. Attempts to produce an all-inclusive registry may add cost but not value, resulting in an overly data collection that reduces quality and completeness.

If the registry has several purposes, these should be translated into specific objectives [81]. This process needs to take into account the interests of researcher, stakeholders, and policy makers. Clear objectives are essential to define data collection and to ensure that the registry addresses the important issues.

The best answers to questions asked by epidemiologists lie in registry. Here below those questions asked when the Pilot Study for the Italian Registry of coronary events was established: 1 - what is the frequency of AMI and who are the persons at particular risk?; 2 - are there any difference in the incidence of AMI between North and South of Italy?; 3 - what proportion of people suffering from AMI recovers satisfactorily and what proportion dies?; 4 - where does that occur, at home or in a hospital?; 5 - is there time for successful intervention by the medical service?; 6 - how long to they wait before calling the doctor? 7 - what is the influence of major coronary heart diseases risk factors on the outcome of the disease?; 8 - are there any environmental factors which may contribute to the development of AMI?; 9 - what happens to patient after leaving hospital?

Time and resources needed to data collection and processing are fundamental [82]. Identification of a core data set is important.

5.1.2 Determine whether the registry is the appropriate tool to achieve the purpose

It should be considered whether a registry is the right tool to meet the set purpose and which data collection is appropriate, between prospective or retrospective data collection. In particular, if data already exist, it must be assessed whether data quality is up to answer the question, whether data are accessible, or whether a new data collection is needed. For example, could the necessary data be extracted from electronic medical records? May the registry avoid re-collecting data? Are data accessible? Is it possible to link administrative data to other relevant data sources?

Once decided that a registry is the appropriate method to collect data, it is important to consider the state of knowledge. Other factors that may influence this decision include the size of the population of interest, how to identify the events, the length of the observational period needed to achieve the objective, and the funds available.

Registries may be the most appropriate choice for some research issues. For example, population-based registries are particularly useful in situations where the occurrence of a chronic disease in the population must be assessed and hospital-based registries are insufficient to address this objective; for example, for CVDs, sudden coronary and cerebrovascular fatal events that do not reach the hospital still represent a significant proportion of overall fatal cases (around 30%); these fatal cases would be lost (not registered) if hospital registries were they only source of data [24, 83]. An estimate of the proportion of fatal cases occurred out of the hospital is of primary importance to plan preventive actions in the general population, since such interventions represent the only way we have to reduce or postpone those fatal cases that do not reach the hospital.

5.1.3 Identify key stakeholders

It is important to consider to whom the research questions are directed. Stakeholders should be identified at an early stage of the registry planning process, as they may have important inputs into the type and scope of data to be collected, may be users of the data, and may contribute in disseminating the results of the registry.

The Ministry of Health, National Public Health Institutes, Regional and Local Health Units, professional and patients associations could be considered as stakeholders; they may be interested in monitoring chronic disease occurrence or community prevention programmes (e.g. agreement with bakery associations for salt reduction in baked products).

The definition of interactions with stakeholders during the planning phase ensures an adequate dialogue and the appropriate sustainability of the registry.

5.1.4 Assess the feasibility of a registry

The key element in the feasibility of a new registry is related to funds and data accessibility. The expenses depend on the scope of the registry, its data collection method, its objectives (research or surveillance), and the data validation process. For population-based registries, systematic direct collection methods (e.g. MONICA) [23] are more expensive and time consuming to implement, compared to a simplified method based on validation of a random sample of events [34-36], or administrative databases and health data routinely collected by national, regional and local health authorities.

Potential sources of funding include Government, Foundations, Health Plan providers, Patient Associations, Private individuals or Entities, Product

manufacturers, Professional societies, Professional societies/pharmaceutical industries, Multiple sponsors.

A public-private partnership is a service that is funded and operated through a partnership (contractual agreement) between a public agency (central or local) and a private entity. There are many good reasons for multiple stakeholders, including government agencies, providers, and industry, to work together for specific purposes; shared funding mechanisms are becoming more common [84].

Information about vital status, death certificate, emigration outside the surveillance area, hospital admission and discharges, dispensed drugs, laboratory tests, co-payment exemptions are usually available at regional level; sometimes this data are not free of charge and accessibility produces additional costs.

5.1.5 Build the team

Several different types of expertise are needed to plan and implement a registry. Some, or all the following types of expertise can be included in the planning of a registry, according to the purpose that the registry is set to achieve; the important thing is to build a team that can work together as a collegial team to accomplish the goals of the registry.

The team should understand the objectives of the registry, its data sources, and the importance of data validation; this is the only way for the registry team to be able to collect and use data in the most appropriate way for the most appropriate interpretation.

Project manager: officer that coordinates the components of the registry to manage timelines, milestones, deliverables, and budgets, ensure communication with stakeholders, oversight committees, and find funding sources.

Experts of the disease/condition under surveillance: they design the registry so that it shall contain the appropriate data to meet its goals and the needs of its stakeholders; these disease experts may be critical to the success of the registry; it is often useful to have on board patient representatives or scientific societies and professional associations. Experts in epidemiology and biostatistics are very important in the design, implementation, and analysis of registry data. Epidemiologists can provide the disease epidemiology and the study design and can work in collaboration with biostatisticians to develop the research objectives and data needed. Researchers can give valuable expertise in health outcomes and health economics.

Data collection and database managers: These experts may need to write specific software to clean data and check data quality, protect, group and store data, and ensure the security of sensitive data.

Statistical analysis and IT-management: epidemiologists and biostatisticians with experience in registry data analysis for the specific disease field are necessary to suggest data to be collected and how to implement the registry. IT personnel shall elaborate the dedicated software to be used to computerise most of the registry's operational activities: record linkage, identification of suspected events, selection of a sample of events to be validated, collection of overall clinic data of the single event to be validated, implementation of the algorithm assessing the diagnostic criteria for event validation, estimation of PPVs, elaboration of main registry indicators and quality control procedures [85].

Legal issues and privacy: Legal and privacy expertise is needed to protect the patients and data owners by ensuring that the registry complies with all ethic and privacy rules (see section 8).

Quality assurance: This is another important component of registry success. Quality assurance will help planning a good registry. Quality assurance goals should be established for each registry, and the efforts made and the results achieved should be described (see section 7.3).

5.1.6 Establish a governance

Governance refers to guidance and decision-making, including purpose, funding, execution, and dissemination of information. Proper governance should provide transparency in operations, decision-making, and reporting of results.

Governance may be assumed by a committee made up of interested individuals who are part of the team (internal governance) or experts who remain external to the work of the registry (external governance).

All governance aspects should be codified in the protocol and manual of operations that can be reviewed, shared, and refined over time. In addition, governance is a dynamic process, and its policies are subject to change with the coming of new evidence that improves the process.

Governance and functions to be considered include:

- *Steering Committee:* it assumes responsibility for major financial, administrative, legal/ethical, and scientific decisions.
- *Scientific Committee:* it includes experts in domains such as database content, general clinical research, epidemiology and biostatistics.

- *Investigators:* registry investigators access and perform registry data analyses to prepare abstracts to submit to scientific meetings, and develop articles for peer-reviewed journals [86].

5.1.7 Define the needed scope and rigor

A) *Scope of data:* some specific variables can characterize the scope of the registry:

- *Population Size:* the population under surveillance should be large enough in order to produce robust indicators to assess time trends and geographical variations (incidence/attack rates, case fatality, survival rate) (See target population 5.1.8)
- *Duration:* it should reflect the length of time during which the registry is expected to collect data in order to achieve its purpose and provide analysis of the data collected. For NCDs (i.e., CVDs, cancer) a long duration (about 10 years) is needed to estimate survival.
- *Setting:* it refers to the specific setting through which the registry will collect data [e.g. administrative databases, clinical records, GPs records].
- *Geography:* population-based registries are established to identify regional differences in incidence rates to better understand the reasons of these differences. The Surveillance Epidemiology and End Results (SEER) Programme of the National Cancer Institute is a population-based registry covering 11 separate geographic areas in the US; it was established in 1972. MONICA coronary registry was established in 37 populations of 21 countries, MONICA cerebrovascular registry was established in 25 cohorts of 11 countries; WHO DIAMOND type 1 diabetes in 114 populations [22]
- *Cost:* budgetary constraints must be carefully considered. Some planning choices, such as building on existing infrastructure and/or linking to current data sources relevant to the purposes of the registry, may increase the return.
- *Clinical data needed:* in some situations, the outcome may be relatively simple to characterize (e.g. death). In other cases, the focus of interest may be a complex set of symptoms and measurements (coronary events) or may require specialized diagnostic testing or tissue sampling (e.g., Computed Tomography-CT, Magnetic Resonance Imaging-MRI in sub typing ischemic/haemorrhagic stroke of cerebrovascular events).

B) *When data needs to be available for analysis:* Meaningful data on disease progression or other long-term outcomes may not be available through a prospective registry for many years, while some indicators need long-term

surveillance (for example, in a cancer registry, a 10-year period of surveillance is needed to produce reliable estimates of 10 year-survival rates; in AMI registry, a follow up period of 1 year is needed, because after 6 months-1 year the probability of a new event is close to the probability of the first event). When possible, and when all needed information is available, retrospective data collection can be implemented. Therefore, the type of data on events and the time of data availability for analysis should be addressed in both short term and long term options.

C) *Scientific rigor*: The content of data to be collected should be driven by the scientific analyses planned for the registry, which are determined by the specific objectives of the registry. A registry primarily designed to monitor disease occurrence will contain data elements different from the elements contained in a registry primarily designed to monitor safety or effectiveness. Similarly, the extent to which data need to be validated will depend on the purpose of the registry and the complexity of the clinical information. For some outcomes, clinical diagnosis (e.g. ICD code from HDRs or mortality) may be sufficient; for others, supporting documents from hospitalizations, referrals, or biopsies may be needed (e.g. coronary events categorized and validated according to standardized diagnostic criteria); whereas others may require a formal adjudication by a committee. Generally speaking, registries undertaken for regulatory decision-making will require increased attention toward validation and diagnostic confirmation (i.e., enhanced scientific rigour).

5.1.8 Define the data set, events, target population, indicators

A) *Core data set*: Data to be included in the registry must be chosen by a team of experts, preferably with inputs from experts in biostatistics and epidemiology. Each variable should relate to the purpose and specific objectives of the registry; each piece of data should address the central questions for which the registry was designed. It is useful to consider the generalizability of the information collected. No superfluous data should be collected. A quality assurance plan (see section 7.3) should be considered when developing the core data set.

The core data set variables (usually MDS) define the information needed to address the questions for which the registry was created. As a minimum requirement, planners must account for these fields when calculating the resources needed and the overall design of the registry. If additional noncore variables (“nice to know”) such as more descriptive variables are included, it is important for such elements to be in line with the goals of the registry and take into account the burden of data collection. A parsimonious use of “nice to know” variables is important for several reasons. When data elements change, there is a cascade effect on all dependent components of the registry process and outputs. For example, the

addition of new data may require changes to the data collection system, training of site personnel on data definitions and collection practices, adjustments to the registry protocol, and amendment submissions to institutional review boards. A registry should avoid attempting to accomplish too many goals.

It is useful to consider which data are already available and/or collected and which additional data need to be collected. When determining additional data, it is imperative to consider whether the information desired is consistent with general practice or whether it might be more intensive or exceeding usual practice. For some purposes, it may be necessary to collect specific laboratory results, or additional visits, or telephone calls, but this could change the way in which the registry is perceived by institutional review boards or ethics committees. Generally speaking, population-based registries are “observational” and not “interventional”. From a regulatory point of view, the “Interventional” element may add significant burden and cost to the registry programme.

Finally, in defining data sets, it is important to consider patient privacy and also national and international ethics rules.

B) Events: The definition of an event is of great importance. Standardised epidemiological definition of events should be adopted as far as possible, taking into account specific operative conventional agreements; e.g.: in the identification of coronary and cerebrovascular events, if a patient experiences further acute symptoms suggestive of an event within 28 days after the onset of a first episode, this second episode is not counted as a new event. Equally, if a patient experiences further acute symptoms suggestive of an event after 28 days from the onset of a first episode, this second episode is counted as an event (recurrent event). It should be noted that each event is registered separately [23].

Some registries generate statistics on incidence (first event); with improving survival a patient can have more than one first event of cancer; this is becoming more common, so a new case must be distinguish from recurrence or metastasis of an existing case [18].

The disease event should be defined in a clear and unequivocal way, specifying diagnostic criteria, patient's characteristics, exclusion criteria from the registry, and the data sources to be used. For example, for a type-1 diabetes registry, the case is defined as: a) diagnosis of insulin-dependent diabetes mellitus according to the WHO classification; b) the date of the first insulin administration should be considered as the date of diagnosis; c) the date of diagnosis must follow the date of activation of the registry; d) patient has to be less than 30 years old at the date of the first insulin administration; e) patient has been resident in the geographical area covered by the registry for at least six months before the diagnosis.

The event identification will also help to guide the data set selection, avoiding the temptation to collect “nice to know” data. An extensive data collection is

sometimes required to properly address potential confounders during data analyses.

Methods to ascertain events should be clearly established; diagnostic criteria, level of data details, and level of data validation should also be addressed. Event ascertainment methods must be considered when evaluating data registry, in particular as concerns the sensitivity indicator (the extent to which the methods identify all events of interest) and the external validity indicator (generalizability to similar populations).

For example, in the case of cancer, coronary, cerebrovascular disease, event identification starts with the record linkage of individual data from HDRs and mortality databases, according to the selected ICD codes reported in hospital discharge diagnoses and in the causes of death; record linkage is necessary to avoid double counting of deaths in hospital for which both HDR and death certificate are available. This implies that, in order to avoid erroneous or multiple identification of events, possible errors in individual identification variables must be removed before implementing the record linkage.

C) Target population: A population-based registry may cover the whole country; where this is not feasible, the population under surveillance would include the residents of a defined geographical and administrative area or region for which population data and vital statistics are routinely collected and easily available and accessible each year. Both urban and rural areas should be under surveillance because differences often exist with regard to exposure to risk factors, treatment of diseases and access to facilities.

It is important that all cases concerning the residents of a defined area are recorded, even when the case occurs outside the person's area of residence (*completeness*). In the same way, all cases treated at hospitals located in a specific area but involving patients who reside outside that area must be excluded. If this is not possible, it is important to provide an estimate of the amount of cases lost and establish whether this creates any change or interference with the validity of the observed rate trends over the years.

It is also important to consider to what extent an area is representative of the whole country (*representativeness*): the area's representativeness shall be assessed as concerns mortality rate, distribution of risk factors (socioeconomic status and health behaviour) and health services (specialised hospitals, GPs).

The population under surveillance should be selected in order to produce sufficiently robust estimates of disease rates, from a statistical point of view, so to establish trends and ensure data comparability. If the population-based registry is not national, it is recommended to select more than one area in order to have a comprehensive picture of the whole country; coordination between the areas is

recommended to ensure comparability. One of the goals for registry data may be the generalization of conclusions drawn for defined populations to be applied to broader ones. This implies that registries must use relatively broad inclusion criteria. As an example, results could be generalised to the overall population if the registry includes the resident population of several geographic areas representative of the country [34,36]. The definition of the target population will depend on many factors (e.g. frequency and occurrence of the disease in the population, data accessibility, scope, and cost).

In establishing the target population, attention should be paid to the access to that population data (e.g. mortality records, HDRs, clinical records). It is important to distinguish the ideal situation from the real one. In this regard, at least the following questions should be considered:

- How common is the disease of interest (e.g., AMI, stroke, type1 diabetes, cancer)?
- Can eligible people be readily identified?
- Are other sources competing for data on the same persons?
- Is care centralized or dispersed (this is of fundamental importance to validate events for which clinical data are needed; if clinical records are spread among a large number of hospitals, collection of information for event validation is more expensive and time consuming than when events are concentrated in one or few hospitals)?
- How mobile is the target population (especially when people move out of the area covered by the registry; all hospitalizations, in and out of the registry area, should be identified for the resident population included in the registry and registered during the surveillance period)?
- Are data sources under quality control? Do they assure time continuity of data collection? Are they geographically homogeneous in the area under surveillance of the population-based registry?

An increased accessibility to the target population, and the completeness of the information needed for the registry, guarantee benefits in terms of enhanced representativeness and statistical power.

The target population should be selected taking the following parameters into account:

- *Age*

The age range depends on the registry aim; for registries covering coronary and cerebrovascular events, such as MONICA, the age range chosen was 35 to 64 years.

The EUROCISS Project suggested a wider age-range, 35 to 74 years, or even up to 84 years for stroke events, considering that more than half of the events occur in patients aged 65 or more; it is recommended to present morbidity and mortality data divided in decennial groups, in particular the age groups 35 to 44, 45 to 54, 55 to 64, 65 to 74, and, if possible, 75 to 84.

If administrative routine data are used, it must be considered that diagnostically valid information tends to be less reliable for patients above 74 years of age.

Age-standardised rates are recommended and the European standard population should be used as reference [31,32].

- *Sex*

Differences in incidence and mortality between men and women are well documented in literature for most diseases. Therefore, it is important that the same high quality data collection methods are applied to both women and men.

- *Population size*

The size of the population under surveillance is determined by the number of events. The number of events is determined by the definition of the event and the event rate in the age groups included. In most cases, the population size has to be determined on the basis of mortality statistics. For example, the age-specific mortality rate of ischemic heart diseases is greater for men than for women. This means that, in order to estimate incidence rates in middle-aged subjects with the same degree of precision, the population observed should be larger for women than for men. Usually, to be eligible to participate in a population-based registry, a minimum amount of 300 deaths per year due to that disease is necessary in the population groups aged 35 to 74 years. This minimum amount has been established to detect a 2% mortality trend decrease in event rates per year. If more than one area is enrolled, the same number should be considered for each single area enrolled [31,32].

- *Patient eligibility*

A patient is considered eligible for inclusion in a population-based registry only if he/she is resident in the area under surveillance, meets the selected criteria and has had an event within the defined surveillance time period.

D) Indicators: Registry outcomes are mainly summarized in indicators that measure the occurrence of the disease in a population; for a chronic disease such as CVD and cancers, the major indicators in a population-based registry are incidence rate, case fatality rate, survival rate, attack rate.

- *Incidence rate* - it is the number of the first events occurred in the population under surveillance;

- *Attack rate* - it is the number of events (first and recurrent events) occurred in the population under surveillance; this indicator is used when the population under surveillance is large and it is not possible to distinguish between first and recurrent event; with improving survival, each individual can have more than one event, so it would be important to identify not only the first but also the recurrent events;
- *Case fatality* - it is the proportion of fatal events;
- *Survival rate* - it is the proportion of patients included in the registry and still alive at different time periods (e.g., 1 month, 6 months, 1 year, 5 years).

5.1.9 Develop a protocol

The protocol documents the objectives of the registry and describes how those objectives will be achieved. The study plan should at least include the registry objective(s), the event (case) definition, the geographical area(s) included in the registry, the target population, the diagnostic criteria adopted for validation, and the data collection procedures. A full protocol will document in detail objectives, design, participant inclusion/exclusion criteria, training, validation procedures, quality methods, outcomes of interest, data to be collected, data collection procedures, governance procedures, and plans to comply with ethical obligations, protect privacy and disseminate results.

In addition, registries may have statistical analysis plans including estimates on the population size needed to produce reliable results, a complete description of the methods used to produce estimates of defined indicators (e.g. how to estimate PPV, how to estimate disease indicators applying PPV).

5.1.10 Develop a project plan

The development of an overall project plan has a critically importance; that's why the registry team has a roadmap to guide all collective efforts. The project plan should include:

- a detailed timeline and management plan schedule, to ensure that the registry and its data deliverables are completed on time;
- a cost plan, to keep registry costs within the budget;
- a quality plan, to describe the procedures to be used to test decisions in the registry building process. Such a plan can help in the early detection of design

errors and in formulating the necessary changes to achieve the plan's scope, and in ensuring that data meets stakeholders' expectations;

- a communication plan that specifies who is responsible and to whom communications should be addressed, as well as the frequency and methods of communications;
- risk identification. Some risks are predictable, and therefore can be assessed in the very early stages of registry planning. It is important to prioritise risks according to their potential impact on the specific objectives, and develop an adequate response. Some predictable risks include:
 - disagreement between stakeholders over the scope of specific tasks;
 - inaccurate cost estimates;
 - delays in the timeline.

6. IMPLEMENTATION

This section describes the step-wise procedures required to implement a population-based registry and that should be consistent with the recommendations reported in chapter 5 (Methods).

The flow chart summarises these procedures (Figure 11).

6.1 *STEP 1: Define target population and routine data*

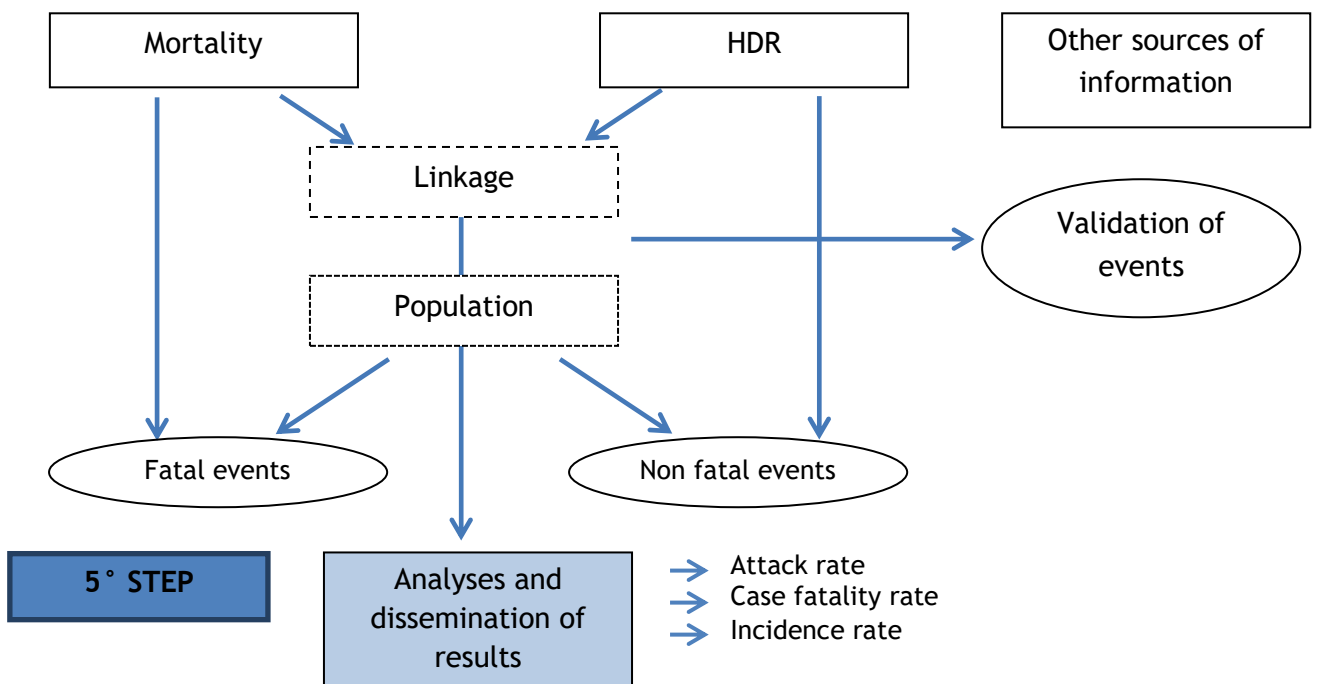
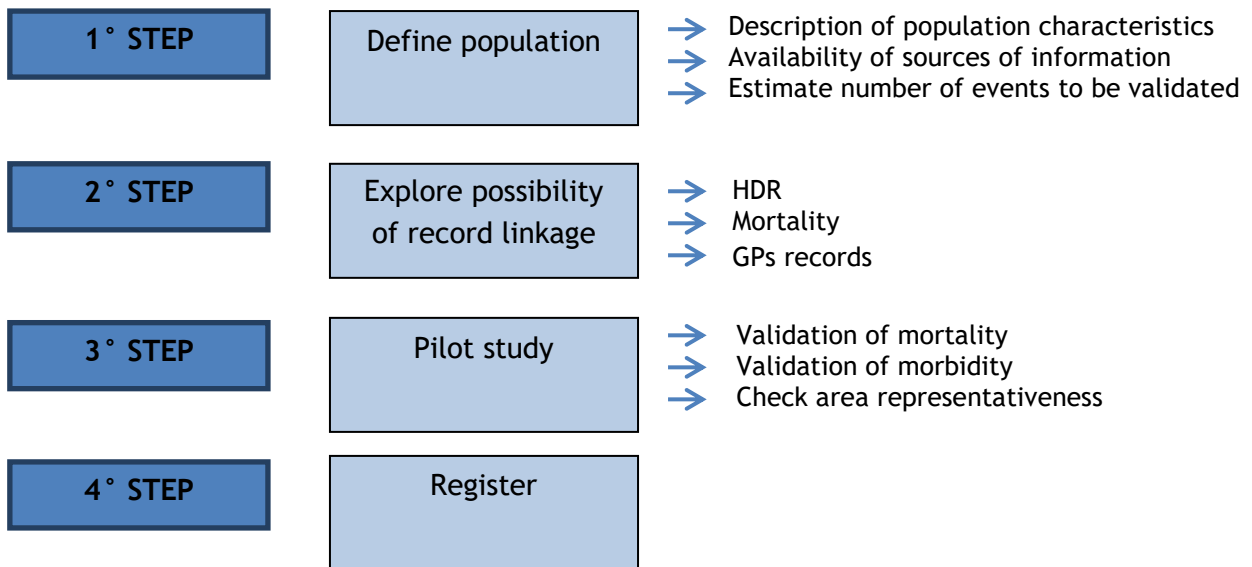
- The geographical administrative area to be selected should have a population big enough to provide stable estimates. This means, for example, that a stable population in a representative area of the country with 300 fatal events in the age range 35 to 74 should be chosen in the case of coronary events under the hypothesis of a 2% annual reduction of attack rates in 10 years [31,32]. If more areas are under surveillance, 300 fatal events should be considered for each separate area.
- The availability of resident population should be checked at mid-year or in inter-census estimates, according to age groups (quinquennial or decennial) and sex. This is important to assess population health indicators such as incidence/attack rate. The numbers of resident population are provided by municipal records.
- A demographic characterisation shall be done for the population under surveillance and shall include a detailed description of this population; in

particular, it should consider the following elements: demographic characteristics (age and gender distribution); socio-cultural characteristics (educational level, occupation, social group, unemployment rate, migration, immigrants with or without citizenship); characteristics of the healthcare system (specialised hospitals, GPs, rehabilitation clinics); macro (North, Centre, South) and micro areas (regions, or urban and rural). Disease frequency is often different in macro areas of the country; a description of differences in mortality and risk factors allows selecting those areas to be included in the surveillance system. In a population-based surveillance, the phenomenon of immigration plays an important role, therefore immigrants coming from European and extra-European countries and resident in the study area should be enrolled. Geographical or administrative borders of the surveillance areas should be clearly defined.

- The availability and accessibility of different data sources must be checked. In some regions, databases are available upon payment; in others, databases are free of charge and accessible upon presentation of a proposal to be submitted and approved by the ethical committee; usually mortality data are accessible at the National Institute of Statistics with a delay of few years, or at the National Public Health Institute. The other databases are available at the Ministry of Health which distributes them on request, in an anonymous form, or at regional level in a detailed form.
- Existing hospital discharge and mortality data and other sources of health care data (e.g., emergency care, specialist visits, diagnostic service, drugs databases) must be analysed. Events occurring in the study area but involving non-residents or non-residents admitted to hospital in the study area do not qualify. Events involving residents of the study area but occurring out of the study area do qualify. Efforts must be made to find them or estimate their potential loss, and whether or not this loss could modify and interfere with the validity of the observed rate trends over the years.
- Problems with these data must be identified: data coverage, ICD version and classification systems of health data, identification of events, procedures, unit of analysis (event or patient), PIN, availability of previously conducted epidemiological studies to assess data coherence. Data files are usually available in detailed form at regional level.

When a register is launched for the first time, it is recommended that future follow-ups are planned to measure trends. This can be achieved by a continuous surveillance, as part of a broader health information system, or through a biannual register implemented at five-year or ten-year intervals.

Figure 11. Description of stepwise procedures.



6.2 STEP 2: Carry out record linkage of administrative data

In the Northern countries, where every citizen has a PIN included in national registries of hospital discharges and deaths, record linkage for the identification of events is efficient and reliable. For countries that have not adopted the PIN, it may be much more difficult to perform this step. Files have to be organised with the same format and have to include the same personal variables needed to univocally identify subjects (family name, first name, date of birth, place of birth, residency).

It is recommended to:

- Explore the feasibility of record linkage within hospital records - probabilistic or deterministic approach based on personal variables or PIN use (within the same hospital, among hospitals of the area under surveillance, among hospitals at regional level). When hospital records are collected at national level, it is possible to include also those non-fatal events occurring out of the surveillance area;
- conduct a propaedeutic but unavoidable activity before implementing record linkage, which consists in accurately checking and cleaning both mortality and HDRs administrative databases from possible errors in identifying variables used for record linkage (family name, first 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 a missing record linkage between corresponding records due to possible errors in identifying variables. These kinds of errors can considerably bias results, since they influence the identification of the first event, the dates of the first and recurrent events, and consequently the number of events for the same subject and for the overall population included in the registry.
- Explore the feasibility of record linkage between hospital records and mortality registers (probabilistic or deterministic approach based on personal variables or PIN use). The *Deterministic* linkage matches records from the two data sets by using a unique variable (e.g. PIN or hospital chart number) or by a full agreement of a set of common variables (e.g. name, gender, birth date). The *Probabilistic* [85] linkage is used to identify and link records from hospital discharges to corresponding mortality records on the basis of a calculated statistical probability for a set of relevant variables (e.g. name, gender, date of birth). This type of linkage puts together records having a specified high matching probability. The method requires a detailed prior knowledge of the various measures to assess the relative importance of specific identification variables in both files to be linked. The main limitations of record linkage are the difficulty in obtaining administrative files for research purposes: mortality data files are usually available at the National Institute of Statistics, whereas

hospital discharge data are available at the Ministry of Health. These kinds of 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 local health units. When combining databases, missing events are mainly explained by PIN or name errors and these errors 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. For hospital discharge diagnosis, the same variables should be considered, with the addition of admission date and hospital discharge diagnoses).

- Explore how to obtain the necessary funds to process large administrative files.
- Explore the feasibility of linkage with other sources of information (e.g. GPs, drug dispensing registers). Not all GPs are organised in networks, with computerised documentation of patient history; when they are, the definition of events is rarely made following the same diagnostic criteria.

6.3 STEP 3: Perform a pilot study and validate routine data

Before starting a registry or a large scale use of linked administrative data, it is recommended to perform a pilot study in a small area on available hospital discharge and mortality data and on other available health care data, in order to study the registry's feasibility and estimate its internal validity.

Validation studies on available data include:

- Coverage estimates: comparison of different routine data sets (electronic or manual), number of patients treated in and out-of-area, hospital/mortality ratios, age and gender ratios, principal vs. secondary diagnoses and/or procedures;
- Validation of discharge diagnoses and every diagnosis reported in health care data sets, according to standard methodological/diagnostic criteria (including revision and abstraction of medical records) in a random sample or in all cases;
- Validation of mortality causes according to standard methodological/diagnostic criteria in a random sample or in all cases; analysis of the demography and representativeness of the area in comparison with the region or country;
- Selection of the age range of interest.

6.4 STEP 4: Set up a population-based registry

Based on Step 2 and 3, it is possible to set up a population-based register following A (record linkage between routine administrative data-based registers) or B (disease-specific data collection).

A. Register based on record linkage between routine administrative data:

- when the linkage procedure between hospital discharge and mortality records is feasible, it is important to define the event and its duration, how to handle transfer between hospitals with difference in the diagnoses between the admitting hospital and the hospital where the patient is transferred, how to define time events, recurrent events, fatal and non-fatal events, etc.;
- validation of diagnostic information is recommended in a sufficient sample size of the identified events randomly selected during a period of the year or during same days each month (in order to trace seasonal variation), with the estimation of sensitivity, specificity and PPV of defined events;
- when the validation of a sample of events is not possible and area-specific PPV cannot be estimated, available PPV, drawn 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;
- population data by age and sex of the area under surveillance are needed to estimate incidence, survival, attack rate, case fatality and mortality rates;
- regular validation should be performed.

B. Register based on disease-specific data collection:

- set up a pilot population-based register with proven standardised protocols and evaluate the pilot study results (coverage, completeness of information and diagnostic validity);
- based on the results of the pilot study, a full scale register should be set up, if feasible, and it should be decided whether to use hot or cold pursuit;
- then, a full-scale register (target population, data collection methods and validation procedures), if feasible, shall be designed.

To set up a full scale register:

- one or more populations representative for the region or the country shall be selected;

- for each selected population, a population-based register with approved standardised protocols shall be set up;
- a detailed protocol shall be written for data collection, and shall include validation procedures;
- information coverage, representativeness and completeness shall be evaluated;
- if relevant, the results from the register shall be used to validate administrative data.

6.5 STEP 5: Analyses and dissemination of results

Registry outcomes are mainly summarized in indicators that measure the occurrence of the disease in a population; for a chronic disease, the major indicators of a population-based registry are:

Incidence rate - it is calculated as the number of first events occurred in the resident population at risk of a defined geographic area during a specific period of time. This indicator can be assessed only if information on first events is available; to be sure of selecting only 'first' events in the calculation of the incidence, and not include 'previous events' occurred before the establishment of the registry, a long period (about 5/10 years) of retrospective observation is needed (e.g. in Northern European countries, an event is defined as first event if in the previous 7 years there is no hospital discharge with the event as primary or secondary diagnosis in HDRs).

Case fatality - it is the proportion of fatal events occurred in the resident population of a defined geographical area during a specific period (e.g., 1 hour, 28 days); all the in and out-of-hospital fatal and non-fatal events should be considered in this denominator.

Survival rate - it is the proportion of patients included in the registry and still alive at different time periods (e.g., 1 month, 6 months, 1 year, 5 years).

Attack rate - it is calculated as the number of events (first and recurrent events) occurred in the resident population of a defined geographical area during a specific period of time.

To plan a fruitful dissemination of results, the following things are needed:

- definition of a strategy to analyse data and disseminate results to decision-makers, stakeholders and the population at large;

- yearly web publishing of incidence/attack rate, case fatality and survival rate indicators, according to gender and age-standardised rates with the European population as reference (35 to 74 and 35 to 84);
- utilisation of data for research. This is very important to ensure a high quality of the registry over time. And high quality registers can be the basis for good research.

7. QUALITY CONTROL

Quality assessment is extremely important for a correct use of registry data.

The term “quality” can be applied to registries to describe the confidence that the registry design, management, and data analysis can offer against bias and errors in inference, which means erroneous conclusions drawn from the registry.

In determining the utility of a registry for research or decision-making, it is critical to understand the quality of the procedures used to obtain the data and the quality of the data stored in the database, which is extremely important also for a valid monitoring and comparison between regions and countries. Data management - i.e. the integrated system for registry data collection, cleaning, storage, monitoring, reviewing and reporting - determines data utility to achieve the goals of the registry. All quality control steps are influenced by registry hypothesis and objectives.

Registry protocols should consider how to ensure quality to a sufficient level for the intended purposes and should also consider how to develop appropriate quality assurance plans. Registry managers should assess and report on those quality assurance activities; data collection and validation procedures have better quality if performed in accordance with international guidelines or standards.

The quality of registry data is evaluated according to data completeness, validity and timeliness.

Quality evaluations of any registry must be done with respect to the registry’s purposes and should consider the disease, the type and the purpose of the registry, its feasibility and affordability [5].

7.1 *Completeness*

Completeness is the extent to which all relevant information to achieve registry objectives is taken. This paragraph presents the principles and methods available to determine the extent to which this ideal is achieved. These methods may be

used to evaluate overall completeness of the registry database, defined, for example, by type of events, area of residence, or age group.

Completeness of a registry has two levels: completeness of events and completeness of information that is relevant to provide final indicators, i.e. the objectives of the registry.

7.1.1 Completeness of event identification

Completeness of events means that all events in the target population have been included. Possible bias derive from difficulties in case identification, such as the need to cover hospitalised cases whenever they occur, during day/night or winter/summer, within the region and outside it, as well as cases occurring outside the hospital (e.g. sudden death among patients who never reach the hospital) when the registration system is based only on hospital records. Another source of potential loss of case identification is private practice: private physicians and clinics may be less cooperative than public ones; in private hospitals, the staff may be more sensitive to criticism and anxious to show how they register medical documents. The identification of fatal events could also be potentially less difficult than that of non-fatal events (death is a definite event, documented by death certificate); whereas survivors may be lost in the totality of the inhabitants of the surveillance area, death is unequivocal. In some registers, case-finding could be more problematical in the elderly, since multiple pathologies may make hospital admission less likely, and extracting diagnosis from hospital information systems (or death certificates) may be more difficult; in other registries, the opposite occurs, since younger patients may have had no other illness episodes and the records may be restricted to just the relevant events. Moreover, the less severe cases do not appear in hospital discharge databases, and their absence presents a barrier to population-based incidence or prevalence studies. It is important to estimate the magnitude of the loss of cases and establish whether it could interfere with or change the validity of the observed rate trends over a period of years.

The identification of potential cases could be based on many different sources, depending on the registry purpose: paper or electronic health records (hospital discharges, death certificates, etc.), administrative databases, other registry information, new information collected from the patient, new or existing information reported by or derived from clinicians and medical records. Another source of bias could be the duplication of events. For example, duplicate cases may include cases transferred from one ward to another, which could be extended to events included twice because more data sources are used.

The completeness of event identification also depends on the existing standard of medical care: if the medical care system misses or misdiagnoses cases, the register cannot remedy the omission.

Data collection procedures need to be carefully considered in planning the operations of a registry. Successful registries depend on the sustainable flowchart used.

7.1.2 Completeness of information

Completeness of information means that all relevant event information has been registered (e.g. place of treatment, date of admission, date of discharge, unique identification number, sex, hospital discharge diagnosis codes, intervention/procedure codes, department/ward, date of birth). The intent of any analysis is to make valid inferences from the data. Information included in the registry may be based on many different types of sources: paper forms, direct data entry, facsimile or scanning systems, interactive voice response systems, and electronic forms. Each type of source has a specific way to contribute to the incompleteness of information. Missing data can be a challenge for any registry-based analysis, including situations where a variable is directly reported as missing or unavailable, a variable is “not reported” (i.e., the observation is blank), the reported data may not be interpretable, or the value has to be imputed as missing because of data inconsistency or out-of-range results.

7.1.3 Evaluation of completeness

It is useful to separate the completeness evaluation methods into two categories: qualitative methods providing an indication of the degree of completeness relative to other registries, or over time; and quantitative methods providing a numerical evaluation of the extent to which all eligible events have been registered.

7.1.4 Qualitative method evaluation

There are several methods that provide some indication of the completeness of a registry, but that do not quantify the number of cases missing. These include the following: 1. historic data methods, such as stability of incidence rates/prevalence over time and in different populations, shape of age-specific curves; 2. mortality-incidence ratios and fatal and non-fatal event ratio; 3. number of sources/notifications per case.

1. As concerns *historic data methods*, most registries systematically review their data for unexpected or implausible trends in incidence, as a potential manifestation of changes in completeness of registration. The concept can be extended to include comparisons of results with those observed in other populations that might have been expected to have similar rates. Differences among different regions may reflect specific local variations in prevalence of risk factors, or the presence or intensity of screening; nevertheless, systematic discrepancies provide evidence of possible under-registration (or over-registration, due, for example, to the inclusion of duplicate records). Age-specific incidence rates could also be examined during the process, in order to detect abnormal fluctuations, including any fall in the rate of increased incidence or prevalence in older subjects (suggestive of under-ascertainment in the oldest age groups). The curves also reveal problems with estimates of population at risk for specific age groups.
2. A *mortality/incidence ratio* is also an important indicator of completeness, an example of the 'independent case ascertainment method'. It is a comparison of the number of deaths, obtained from a source independent from the registry (usually, the vital statistics system), and the number of new cases of a specific disease registered, in the same period of time. Mortality/incidence ratio values greater than expected lead to a suspicion of incompleteness (i.e. cancers or cardiovascular events missed by the registry).

Also the fatal and non-fatal event ratio is an important indicator of completeness. It is a comparison of the number of fatal and non-fatal events obtained from independent sources of information in the same period of time. For example, fatal coronary events are approximately 1/3 of total events, as shown in MONICA study [24]. If the proportion of non-fatal events is not expected, it may indicate incomplete registration of mortality data or, more likely, of non-fatal events. Since both survival and quality of mortality statistics are related to the level of socio-economic development, the geographical area under surveillance is important in evaluating the indicators.

3. For some registries, another indicator of completeness is the *number of sources/notifications* per event. The rationale for using as many sources as possible is that this reduces the possibility of disease diagnoses going unreported, thus increasing the completeness of registry data. Two indices have been used as indirect indicators of completeness: the average number of sources per event, and the average number of notifications per event. Efficient record linkage is essential to recognise a same case notified several times.

7.1.5 Quantitative method evaluation

Three methods are available to obtain a quantitative evaluation of the degree of registration completeness: 1. independent case ascertainment; 2. capture-recapture method; 3. number of missing values per variables.

1. *Independent case ascertainment* : it includes a re-screening of the sources used by the registry (case-finding audits), to detect any event missed during the registration process (i.e. unmatched cases), the use of one or more independent sources of events, and their comparison with the registry database. The first method involves independent re-ascertainment of records, usually in a sample of hospitals/facilities and, in each facility, in a sample of time periods. Records of events identified during the audit are enumerated and matched against the registry's files. Unmatched events are followed back to verify their reportability, and the percentage of actually missed events that should have been reported is then calculated. Most registries using this method focus on hospital reporting and thus provide an estimate of the completeness of reporting for hospitals only, not a true estimate of completeness for a multi-source population-based registry.

Comparison of the registry database with sets of events identified independently of the registry's case-finding procedures is a particularly useful and objective method to evaluate completeness. It requires record linkage between the registry database and the independent sources of events, to estimate the numbers of events missed by the registry. Independent sources could include international clinical follow-up studies, multi-centre clinical trials, the database of a GP network, cohort studies, multi-hospital case-control studies, hospital databases not accessed by the registry, or community surveys.

2. *Capture-recapture method*: it has been also advocated for use in estimating completeness of disease registers. The capture-recapture method was originally developed to estimate the size of a closed animal population. Its procedure captures, tags and releases as many animals as possible in a specific area and at a specified time - the 'capture' stage. At a later time, this capture is repeated - the 'recapture' stage. The numbers of animals in each sample and the numbers of animals common to both samples are used to estimate the numbers in the total population (assuming that capture and recapture are independent). Practically, capture-recapture analysis of completeness requires that record linkage is successfully carried out (so that events identified by each of the multiple sources are correctly classified), and that, if death certificates are used as a source, the relevant death is correctly identified on them. In addition, two assumptions are made when

using the simple capture-recapture method. The first is that when there are two sources, the identification (capture) of a case by one of them is independent of the other, and, more generally, there is no dependency between all sources in a multi-source model. The second is that all individuals have the same probability of being captured. Neither can be directly tested, and violation of either could lead to over or under-estimation of the true patient population size [87].

3. *Number of missing values per variables* : it is also important to keep in mind that the impact of data completeness will differ, depending on the extent of missing data and the intended use of the registry. It may be less problematic with regard to descriptive research than with regard to research intended to support decision making. For all registries, it is important to have a strategy for the identification and handling of missing data, as well as the explicitly reporting on data completeness to facilitate the interpretation of study results. The number of missing values per each registry variables could be also assessed as an indicator of registry completeness.

Finally, all aspects considered, the degree of data completeness should be summarized and made available to researchers and possible users of registry analyses.

7.2 Validity

Validity refers to the proportion of register events and exposures having a certain characteristic (e.g. definite AMI); the greater is the ability to grasp the proportion of people who actually had the event (internal validity) and generalize these results to the target population (external validity), the better is the validity of the registry. Validity depends on the study design (how to validate events), the precision of source documents and the level of expertise in abstracting, coding and recoding data. Validity assessments answer the question: does data match the rules?

Population-based registries that observe real-world settings may collect all of the information needed to assess patient outcomes in a generalized way, but interpreting this information correctly requires analytic methodology geared to address the potential sources of bias that challenge observational studies. Interpreting registry data also requires checks of internal and external validity and sometimes the use of external data sources to validate key assumptions (e.g. number of coronary events should be coherent with the risk factor distribution in the general population and with mortality rates for IHD).

Data quality is dependent on the entire chain of data collection, identification, validation and processing.

7.2.1 Internal validity

The potential for bias refers to opportunities for systematic errors to influence the results. Internal validity is the extent to which results are free from bias, and the reported association between exposure and outcome is not due to unmeasured or uncontrolled-for variables. These attributes include the extent to which a registry identifies all events of interest and the extent of accuracy of exposure data, to support the main research questions.

Poor internal validity could particularly affect population-based registries aimed at assessing incidence and survival rates that use healthcare databases not collected for research (e.g. administrative databases). These registries require special considerations because they may produce biased or invalid results; the challenges faced by healthcare databases include inaccurate measurement of exposures, outcomes, confounders and overweighting of results. Discharge diagnoses, for example, are not always validated on a routine basis, and validation studies are necessary in all countries to check diagnostic quality and data comparability. The validity of a hospital discharge diagnosis might vary on the basis of patient characteristics, geographical region, and type of hospital or clinic. Moreover, hospital admission policies may vary over time and place.

The internal validity of a registry is strongly influenced by its ability to collect diagnostic information and correctly identify events, and by its ability to implement an accurate exposure collection. The diagnostic information for a given event is valid if it measures the disease it claims to measure. Potential sources of error relating to accuracy and falsification should be rigorously evaluated and quantified to the extent feasible.

The assessment of a registry's **internal validity** is usually done through the evaluation of its sensitivity and specificity, as well as through the calculation of the predictive value of the registered diagnosis compared to a gold standard. For example, to validate coronary events, the MONICA diagnostic criteria, the new criteria of the Joint ESC/ACC, the AHA criteria may be applied as gold standards [23, 26-28]. Validation studies of routine statistics have been carried out over the years with heterogeneous results due to differences in methodology or true differences in the validity of the routinely collected data between countries [73-76]. Some studies have been conducted, which compare community registers, national statistics and data from epidemiological projects [73,75]. These findings stress the importance of validating routine mortality and hospital statistics to determine whether and how they can be used to reflect true attack and mortality

rates. Consistency of coding with the diagnosis and consistency of coding/comparability of the information for different areas of the country and over time represent other issues for validation. As mentioned, registry manuals should provide a clear, unambiguous definition of the outcome being studied, a description of how it will be measured, and a discussion of the accuracy of that measurement.

Using the diagnostic criteria, it is possible to evaluate whether the tools used to establish the application of valid methods are different for hot pursuit and cold pursuit.

The validation of the diagnostic information recorded in the register can include the examination of either all events or just random samples. If it is not possible to validate all events included in the disease register or in the mortality routine statistics, the validation objective should be to evaluate a sample of events. The sample should be distributed along a full year in order to ensure that potential seasonal or other time-related variations of diagnostic patterns are traced. Relevant data must be checked periodically by sampling, as it is usually not feasible to check all data. For local registries with a limited number of events, it might be possible to validate each single event, whereas national registries (for practical reasons) can only validate data on the basis of random samples of suspected events recorded during a selected period or during some days each month. One sampling method consists in choosing some days each month and evaluating all the cases that occurred in those days, extracted from either hospital discharge records or mortality records, applying diagnostic criteria. In this way, seasonal variation can be traced. For registers based on hospital discharge and death certificates, it is necessary to review and validate the diagnosis in at least a sample of events, against the gold standard chosen, in particular when they are followed by a secondary cause in the discharge or death certificate. Some countries only code the principal cause of death, while others code all secondary causes of death. Those who rely on principal cause of death only should perform validation at least twice every 10 years and for a full year or on a sufficiently sized sample for a full year. Depending on the percentage of false negative diagnoses for the event or death found in the first validation, decisions should be taken about the intensity and duration of the validation exercise for fatal cases throughout the registration period. A false negative rate above 10-15% should in principle be an indication to perform diagnostic validation of death certificates on a continuous basis rather than on a periodic or sample basis.

7.2.2 External validity

External validity, also known as generalizability, is a concept that refers to the utility of the inferences for the broader population that the registry is intended to represent. One of the goals for registry data may be to enable generalize the conclusions obtained for a defined populations to broader populations, and

therefore the inclusion criteria for most registries are relatively broad. No particular method can ensure that an approach to patient recruitment is adequate, but it is worthwhile to note that the way in which patients are recruited, classified, and followed can either enhance or diminish the external validity of a registry.

Here below there are some examples of how these methods of patient recruitment and follow-up can avoid systematic errors:

- It is not essential that the whole country is covered by a surveillance system, but it is essential that event registration is complete with regard to events occurring in the population under surveillance.

- It is important to know how representative the registry is for the whole country according to the mortality rate, the distribution of risk factors (socioeconomic status and health behaviour) and the distribution of health services (specialized hospitals, GPs).

- A careful description of the population characteristics may help to define how representative the population under surveillance is for the whole country.

The strong external validity of registries is achieved by the fact that they include general population. Therefore, registry data can support observational studies providing a good description of the course of disease and impact of interventions in current practice and, for some purposes, may be more relevant for decision-making than data derived from a clinical trial. In fact, even though registries have more opportunities to introduce biases (systematic errors) because of their non-experimental methodology, well designed observational studies can approximate the effects of interventions observed in randomized clinical trials.

As regards the external validity of a registry, results are mainly affected by the representativeness of the registry target population with regard to the range of characteristics reflective of the broader population that is the object of epidemiological registry-based studies.

Mainly, representativeness should be considered in terms of patients (e.g. men and women, children, elderly people, different racial or ethnic groups). Involved patients should have demographic characteristics and exposures similar to those of the population target, on which inferences will be done.

Registry study designs often restrict access eligibility to individuals with certain characteristics (e.g. age) to ensure that the registry will have subgroups with numbers of patients sufficient for the analysis. These criteria are also used to exclude people with characteristics that can introduce significant biases into the estimates of disease frequency or estimates of association, and that cannot be easily or adequately controlled in the analysis.

The representativeness of the various sites (healthcare providers, hospitals, etc.), is often considered in terms of geography, practice size, and academic or private practice type. A revision and redefinition of research issues can help researchers define an appropriate target population and a realistic strategy for subject selection.

Two items can be reported to help the user and assess the generalizability of research results based on registry data: a description of the criteria used to select registry sites, and the characteristics of these sites, in particular those characteristics that might have an impact on the purpose of the registry.

The assessment of registry overall validity may be also done through comparisons with external information. Examples include the rates, or prevalence, of the outcomes of interest in other studies and different data sources (taking into account reasons why they may be similar or different). Such comparisons can put the findings of registry analyses within the context of previous study results and other pertinent clinical and biological considerations as to the validity and generalizability of the results. An external comparison group could be a group of patients similar to those who are the focus of interest, but who do not have the condition or exposure of interest, and for whom relevant data that has been collected outside of the registry is available. External comparison groups can provide informative benchmarks to understand the observed effects, as well as to assess generalizability. Additionally, large clinical and administrative databases can provide useful information on comparable subjects at a relatively low cost. A drawback of external comparison groups is that data is generally not collected in the same way, and the same information may not be available. The underlying populations may also be different from the registry population. In addition, plans to merge data from other databases require the proper privacy safeguards to comply with legal requirements for patient data.

7.3 Assurance of data quality

Quality assurance aims at affirming that data was collected in accordance with established procedures and that data meets the requested quality standards to accomplish the registry intended purposes and the intended use of data. Quality assurance activities generally fall under three main categories: 1. Data quality assurance, 2. quality assurance of registry procedures, and 3. quality assurance of computerized systems. Since many registries are large, the level of quality assurance that can be obtained may be limited by budgetary constraints.

1. Requirements for **quality assurance should be defined during the registry planning**. Because certain requirements may have significant cost implications, a quality assurance plan is recommended. It should be based

on **identifying the most important sources of error or potential bias in procedures** that may affect the quality of the registry in the context of its intended purpose.

Clarity of description and standardization of definitions are essential to data quality assurance. As described above, data collection procedures for each registry should be clearly defined and described in the detailed manual of operations. Event inclusion and exclusion criteria and the validation procedures should be reported. A data dictionary should also be done, with a detailed description of each variable used by the registry, including the source of the information, coding information and normal ranges. Data definitions include ranges and acceptable values for each individual data. For example, logic checks for data capture validity may be created for variables that should be mutually exclusive. Data managers should develop formal data review manuals to be used by reviewers and data entry personnel. Registries to evaluate events should use predetermined and defined uniform and systematic methods of data collection, all data-related procedures—including the sources of data, data elements and their definitions and data quality requirements.

2. Generally, **registry manuals of operations** includes instructions to search **data** that will go into the registry (e.g. specific diagnoses or laboratory results). **Data codification** could include either standardized codes from a codebook (e.g. the ICD-10 code) corresponding to a text diagnosis in a chart, or codes that may be unique to the registry (e.g. a scale from 1 to 4 for definite, possible, insufficient data, non-event). All abstraction and coding instructions must be carefully documented and incorporated into a data dictionary for the registry. Because of the “noise” in unstructured, hard-copy documents (e.g. spurious marks or illegible writing) and the lack of precision in natural language, clinical data processed from the same documents by different abstracters may differ. This is a potential source of error in a registry. Manuals should also include data validation rules referring to the logical checks on data entered into the database against predefined rules for either value ranges or logical consistency, with respect to other data fields for the same event.

Data validation rules should also include information on how to handle missing data; invalid entries (e.g. multiple selections in a single-choice field, alphabetic data in a numeric field); erroneous entries (e.g. patients of the wrong gender answering gender-based questions); and inconsistent data (e.g. an answer to one question contradicting the answer to another one). Guidelines should consider procedures to remedy these data problems. For example, a data error on an interview form could require to query the interviewer or the patient, or to refer to other data sources that may be

able to solve the problem. Documentation of any data review activity and remediation efforts, including dates, times, and results of the query, should be maintained.

3. **Data cleaning** refers to the correction of data problems, including missing values, incorrect or out-of-range values, responses that are logically inconsistent with other responses in the database, and duplicate patient records. How and to what level data will be cleaned should be addressed upfront in a data manual identifying data elements to be cleaned, describing data validation rules or logical checks for out-of-range values, explaining how to handle missing values and values that are logically inconsistent, and discussing how to identify and manage duplicate patient records. Ideally, automated data checks are pre-programmed into the database for presentation at the time of data entry. These data checks are particularly useful for cleaning data at the site level, when the patient or medical record is readily accessible. Even relatively simple edit checks, such as range values for laboratories, can have a significant effect on improving data quality. The automated mechanisms of numeric checks and alerts can improve validity and reliability of data collected. Data managers shall carefully review the data, using both data extracts analysed by algorithms and hand review, to identify discrepancies and generate “queries” to send to the sites for solution. It is very difficult to foresee all potential data discrepancies at the time of development of the data management manual and edit checks. Therefore, even with the use of automated data validation parameters, some manual cleaning is often still performed.

The creation of explicit **data definitions for each variable to be collected is essential to select data elements**. This is important to ensure the internal validity of the proposed study, so that all participants in the data collection acquire the requested information in the same reproducible way. When deciding on data definitions, it is important to determine which data elements are required and which elements may be optional. This is particularly true in cases where the registry may collect a few additional “nice to know” data elements. The determination will differ depending on whether the registry uses existing medical record documentation to obtain a particular data element or whether the clinician is asked directly. However, if clinicians are asked to provide this information prospectively, they can immediately do it. Moreover, accounting for missing or unknown data should be taken into consideration. In some cases, a data element may be unknown or not documented for a particular patient, and a follow-up with the patient to answer the question may not be possible. Therefore, the form should contain an option for “not documented” or “unknown” data, so that the person filling in the case report form could provide a response to each question rather than leaving some unanswered. Depending on the analysis

plans for the registry, the distinction between undocumented data and missing data may be important.

As regards reduction of **false-positive cases**, strategies can foresee that some evidence in the patient record of medical procedures (e.g. cholecystectomy for gallstone disease or podiatry examination for type 1 diabetes) or interventions (e.g. insulin or glucose-lowering medications for type 1 diabetes) could provide greater confidence in the validity of the event definition. Such an approach often results in a reduced number of cases included and a reduced precision, but provides improved validity. **Registration of causes of death may be incorrect** and may need to be validated, and the collection of information on deaths occurring outside the area of residence has to be ensured. It is to be expected that some events occur outside hospital. **Duplicate registration** of the same case should also be avoided by paying careful attention to record linkage during the registration process. When the event is defined (codes and duration), it may be easy to identify duplicate coding and take out information for quality control purposes. Duplicate codes may include events transferred from one ward to another. In some cases the duration of the admission is very short (<2 days) either because the patient is transferred elsewhere or because of diagnosis misclassification. These cases may also be picked up for validation.

As regards accuracy improvement of data exposure, one frequent shortcoming of epidemiologic research is to compare incidence of disease in an index group with the incidence of disease in all other groups who do not satisfy the index group definition. Such methods are easily applicable in administrative databases, due to the abundance of participants who do not meet the index group definition. This “all others” reference group is therefore usually a poorly defined mixture of individuals. For example, if a pharmaceutical registry is used to compare the incidence of a disease in statin users with the incidence of disease in those who do not use statins, the nonusers’ reference group will contain individuals with indications for statin use but who have not been prescribed statins, as well as individuals without indications for statin use. Nonusers also differ from users in the frequency of contact with medical providers, and this raises the potential for differential accuracy ascertainment of health outcomes. It is therefore preferable to first ensure that the nonusers’ reference group contains individuals who have indications for treatment, and who, if possible, receive alternative therapies for the same indication. If the different categories of statins - such as hydrophilic and hydrophobic statins - are assigned to patients according to the patients’ biological characteristics, then a comparison between users of hydrophilic statins and users of hydrophobic statins is often more valid. On the basis of these definitions, only individuals with indications for statins, and treated with statins, are included in the analysis, and therefore the possibility of confounding is reduced in comparison to a follow up carried out on other indications.

It is always important to clearly define the registry objective, the patient population, as well as potential confounders and modifiers. Researchers must also understand the conditions under which data were originally collected. The selection of data elements requires balancing factors such as their importance for registry integrity and for the analysis of primary outcomes, their reliability, their contribution to the overall burden for respondents, and the incremental costs associated with their collection. Selection begins with the identification of relevant domains. Specific data elements are then selected, with a focus on established clinical data standards, common data definitions, and the use of patient identifiers. It is important to determine which elements are absolutely necessary, and which ones are desirable but not essential. **Overall, the choice of data elements should be guided by parsimony, validity, and a focus on achieving the registry purpose.** Information on behavioural and lifestyle factors (e.g. tobacco use, alcohol drinking, exercise habits, and diet) is infrequently captured or is poorly measured in many databases. Some databases can provide proxy measurements of these behavioural factors. For example, poor lung function or diagnosis of chronic obstructive pulmonary disease is a proxy marker for tobacco smoking history; alcohol-related diseases, such as cirrhosis, or prescriptions for disulfiram can be used as proxy markers for alcohol abuse, and medically diagnosed obesity may be a proxy marker for poor diet and lack of exercise. However, none of these proxies provides a reliable measure of the actual concept.

To prevent the low validity of a registry, and, in particular, the loss of generalizability, it is important to consider to what extent an area is representative of the whole country (representativeness): it could be representative according to mortality rates, the distribution of risk factors (socioeconomic status and health behaviour), and the distribution of health services (specialized hospital and GPs). In some countries, it might be better to start implementing a register with high-risk areas. The population to be monitored should be selected to produce estimates of disease rates that are sufficiently robust from a statistical point of view, so that trends can be established and data comparability ensured. In general, it is necessary to select more than one area representative of socioeconomic or ethnic differences, so to have a comprehensive picture for the whole country, and it is recommended to establish a coordinating body between the areas to ensure comparability.

8. ETHICAL ISSUES

Population-based registries for epidemiological studies contain information about the health status of a person; this is considered sensitive data by law and therefore subject to protection. Moreover, in our society, it is necessary to keep personal

records for different reasons (social, health-related and civil ones), and perform a record linkage between different sources of data. The identification of each subject in the different situations (registries or use of administrative data) has generated legal and ethical problems, given that the personal integrity (autonomy, confidentiality and privacy) of individuals has to be guaranteed. In this section of the manual, only a few aspects regarding ethics and law requirements will be taken into consideration. In particular, consideration shall be paid to the fundamental principles applied through norms, and the latest EU Regulation (2016/679) on the protection of natural persons.

Specific arguments about the topic will be treated by experts in the Horizontal Activity-Ethical and legal issues in Health Information and in the WP10 and WP11.

8.1 *Fundamental principles*

Attention must be paid to the following principles, when considering the ethical and legal aspects of providing information to a population-based registry and when these data elements are used for the legitimate purposes of the registry or when individuals want to exert their right of withdrawal (and therefore ask for destruction or anonymity of their personal data). The principle of “*respect*” of persons generates other ethical aspects, such as the protection of privacy and dignity, confidentiality of the information, and minimization of potential harms. Privacy refers to people, while confidentiality has to do with how the information is handled. Consequently, “*privacy protection*” addresses the right of the subject to decide and control what information about himself is disclosed, to whom and why. “*Confidentiality*” is the right of every subject whose private data is disclosed to be ensured that the person receiving such private information does not communicate it to third parties, unless authorized by the subject himself.

It is necessary to take into account the historical evolution of international documents when considering the ethical principles pertaining to biomedical and epidemiological research, and, in particular, the impact on individuals’ rights when these activities put at stake personal autonomy, freedom of choice, self-determination and privacy protection. Among these documents, we find the Declaration of Helsinki, a set of ethical principles issued by the World Medical Association regarding human experimentation (2013, “Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects”). This is not a legally binding instrument under the international law, but draws its authority from the extent of its ratification and transposition into the national legislations. Anyway, it has to be considered an achievement for all humanity. However, the future of this Declaration raises conflicting opinions, as it has undergone frequent revisions.

While the Helsinki Declaration is object of discussions that could undermine its authority in the long run, another important document is universally recognized as a general and universal set of norms in the field of research ethics. This document is the Belmont Report (full title: The Belmont Report “Ethical Principles and Guidelines for the Protection of Human Subjects of Research, Report of the National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research”); it set an important goal in the evolution of ethics and represents a benchmark for the creation of guidelines in biomedical research worldwide [87].

The Belmont Report, issued on 30 September 1978, is a summary of the basic principles and guidelines developed to assist in resolving ethical problems when conducting research on human subjects. Elaborated by the National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research in USA [Department of Health, Education and Welfare (DHEW), Washington, DC: United States Government Printing Office], it identifies three fundamental principles: respect for persons, beneficence, justice.

Respect for persons - everyone must be treated as autonomous beings responsible for themselves (self-determination), able to take decisions for themselves. In case of persons not able to take decisions for themselves (children, sick people), protection from coercion by others has to be considered.

In the case of population-based registries, the above mentioned principle is well represented and supported by the “*informed consent*” of the subjects to the use of their HI for research purposes. A general ethical requirement for consent clearly implies that human subjects voluntarily permit the use of their HI in a registry, unless a specific exception to voluntary participation applies to the registry itself. One such exception is a legally mandated, public health justification for the compilation of HI (e.g. certain infectious disease reporting). Voluntary agreement to the use of HI in a registry necessarily allows a subsequent decision to discontinue participation.

The consent given for a registry concerns, mainly, two different aspects:

- Consent to create the registry by compilation of patient information
- Consent to the use of data for the purpose of the registry and for other declared purposes.

The purpose of the registries should be clearly described in the consent process.

In a voluntary agreement, the decision to discontinue participation is allowed. The consent process should indicate the procedure for withdrawal of information from the registry, and specify if there is a limitation of withdrawal.

Beneficence (do good, do no harm, protect from harm). The Belmont Report states that, "Persons are treated in an ethical manner not only by respecting their decisions and protecting them from harm, but also by making efforts to secure their well-being (...)". When applying this principle to disease and HI registries, registry developers are ethically obliged to minimize potential harms to the individuals or groups whose HI is included in the registry. Risks for privacy and dignity can be contained if the confidentiality is controlled through some technical and administrative safeguards, such as, for example, control on access to registry data concerning personal identifiers. Another way to reduce the risk is to determine the information necessary and respect this limitation. Certain populations of patients may be vulnerable to social, economic, or psychological harm as a result of a stigmatizing health condition. Pregnant women, neonates and adolescents are particular categories that can require additional safeguards. Developers of registries compiling this HI must make special efforts to protect the identities of the human subjects that contribute data to the registry. Beneficence implies the assurance of a valid scientific purpose, when human subjects are involved, as without a valid scientific purpose this beneficence is unethical. To reach this aim, a registry should have a design, some written operative procedures, methodologies and data elements.

Justice (fairness, equitable distribution of benefits and burdens, equal treatment); this principle concerns the recognition of any potential risks to those who contribute HI to a registry, and the probable lack of benefit to those individuals (except when registries are specifically constructed to provide benefit to those individuals). The imbalance of burden and benefit to individuals reinforces the need to minimize the risks from registry use of HI. Precise and well-developed scientific reasons for inclusion (or exclusion) of defined HI in a registry help ensure that the burden placed on individuals as a result of their participation is fair and equitable.

All these ethical concerns can be applied to both research activity and use of HI for non-research purposes. A registry developer must consider all these aspects, as well as the confidentiality and concerns about the identity of healthcare providers. Developers of a public health registry must consider the protection of the identity of the professional service and institutions and, at the same time, allow disclosures in compliance with the law.

In conclusion, some rules must be respected to correct apply these three ethical principles. **The respect for persons can be assured through the informed consent, beneficence can be reached through the assessment of risks and benefits, and finally justice can be pursued by defining the exact profile of registry participant eligibility, using fair and transparent procedures.**

Transparency is a crucial aspect for registries. It is desirable that information about registry operations be public and readily accessible to anyone who is

interested in it. This can help realise the potential benefits of HI-based research. Registry transparency can also help to understand the scientific processes that the registry underlies. Transparency contributes to public and professional confidence in scientific integrity and validity of the registry processes. Public information about registry operations may also increase the scientific utility of registry data and promote interest among other researchers. Registry developers can promote transparency by making available the registry scientific objectives, governance, eligibility criteria, sampling and recruitment strategies, general operating protocol, and sources of data to anyone interested in them. It is important to stress that **transparency and access to information are to be encouraged**. Funding agencies and healthcare providers, however, have an important stake in maintaining public confidence in how HI is managed. The extent of registry transparency should be prospectively negotiated with these entities. A good method to achieve transparency is the **creation of a web site** informing about registry objectives and operations. Ideally, registry information should be available on different media.

8.2 European regulation and data protection

The EC, the European Parliament and the European Council decided to repeal Directive 95/46/EC and unify and strengthen data protection for individuals within the EU in a unique law called “The General Data Protection Regulation (GDPR) on the protection of natural persons with regard to the processing of personal data and the free movement of such data”. The Regulation was adopted on 27 April 2016, and will enter into application on 25 May 2018, after a two-year transition period. Unlike Directives, a regulation does not require any implementing legislation to be passed by national governments. The impact that the new Regulation will have on each national setting will depend on the legal provisions already operative in each country. When the GDPR takes effect, it will replace the current data protection directive 95/46/EC, in force since 1995. **The GDPR was necessary to cover important aspects, such as globalization, technological development, social networks and cloud computing that were not present in the previous law.** The primary objective of the GDPR is to return citizens the control of their personal data and simplify the regulatory environment for international business by unifying the Regulation within the EU. Here below, a glossary with some definitions (extracted from article 4 of the GDPR) is reported, which may help the reader to understand this Regulation. As a whole, the new Regulation simplifies the procedures among the various EU countries. However, it is important to note that, in many European countries, legal regulations regarding the treatment of personal and sensitive data in biomedical and epidemiological fields are already well established and guarantee a high level of privacy protection

and confidentiality. This will certainly provide the GDPR a good framework for the implementation of its provisions [15].

8.2.1 Glossary

Consent of the data subject means any freely given, specific, informed and unambiguous indication of the person's wishes, with which the data subject indicates his/her agreement to the processing of his/her personal data, by means of a statement or a clear affirmative action.

Controller means the natural or legal person, public authority, agency or other body who, alone or jointly with others, determines the purposes and means of personal data processing; where the purposes and means of such processing are determined by the EU or Member State law, the controller or the specific criteria for its nomination may be provided for by the EU or Member State law.

Data concerning health means personal data related to the physical or mental health of a natural person, including the provision of healthcare services, which reveal information about his or her health status;

Personal data means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is the one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.

Processing means any operation, or set of operations, performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

Profiling means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.

Pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

Recipient means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not. However, the public authorities that may receive personal data specified in the 4.5.2016 L 119/33 Official Journal of the European Union within the framework of a particular inquiry carried out in accordance with Union or Member State law shall not be regarded as recipients; the processing of those data by those public authorities shall be in compliance with the data protection rules applicable to the purposes of the processing.

Supervisory authority concerned means a supervisory authority concerned by the processing of personal data; *Supervisory authority* means an independent public authority established by a Member State pursuant to Article 51.

Sensitive data

Third party means a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data.

Here below, we report a list of the GDPR **most relevant articles on the use of personal data for research and statistical purposes** (see in particular: Art. 17, 21, and 89) [15].

Chapter I, Article 1 refers to the **“Subject-matter and objectives”**

Point 1 states: *“This Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data”*.

Point 2 states: *“This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data”*.

Chapter I, Article 2 refers to the **“Material scope”**

Point 1 states: *“This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system”*.

Point 2 indicates all the conditions for which this Regulation is not applicable.

Chapter I, Article 3 refers to the **“Territorial scope”**

Point 1 states: *“This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union or not”*.

Chapter II, Article 5 refers to the “Principles”

Here below there is a list of the fundamental principles governing data processing. Data should respect: a) lawfulness, fairness and transparency, b) purpose limitation, c) data minimisation, d) accuracy, e) storage limitation, f) integrity and confidentiality.

In detail, Point 1 states: *Personal data shall be:*

(a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');

(b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');

(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');

(d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');

(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');

(f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

Chapter II, Article 6 refers to “**Lawfulness of processing**” and indicates the way to perform a lawful processing. It underlines the use of consent.

In detail, Point 1 states: *Processing shall be lawful only if and to the extent that at least one of the following applies:*

(a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

(b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

(c) processing is necessary for compliance with a legal obligation to which the controller is subject;

(d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;

(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Chapter II, Article 7 refers to the “**Conditions for consent**” and underlines that, in case of a written declaration, this has to be clear, understandable, and can be withdrawn. In detail, it states:

1. Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.

2. If the data subject’s consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.

3. The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal.

Chapter II, Article 8 refers to the “**Conditions applicable to child’s consent in relation to information society services**”.

Chapter II, Article 9 refers to the “**Processing of special categories of personal data**”. The processing of this category of data should be prohibited, except for some situations listed in paragraph 2.

In detail, paragraph 1 states: *Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.*

While paragraph 2 states: *Paragraph 1 shall not apply if one of the following applies:*

(c) processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent;

g) processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject;

(h) processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3;

(i) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of healthcare and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy; 4.5.2016 L 119/38 Official Journal of the European Union EN

As stated in paragraph 4, “Member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health”.

Chapter III is dedicated to the “**Rights of the data subject**”

Section 1 concerns Transparency and modalities. Article 12 refers to: Transparent information, communication and modalities for the exercise of the rights of the data subject.

Point 1 states: *The controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 and any communication under Articles 15 to 22 and 34 relating to processing to the data subject in a concise,*

transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child. The information shall be provided in writing, or by other means, including, where appropriate, by electronic means. When requested by the data subject, the information may be provided orally, provided that the identity of the data subject is proven by other means.

Section 2 refers to Information and access to personal data.

Articles 13 and 14, describe, respectively, the following situations: “Information to be provided where personal data are collected from the data subject” and “Information to be provided where personal data have not been obtained from the data subject”.

Article 15 refers to “Right of access by the data subject”.

Paragraph 1 states: *The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information*

(a) the purposes of the processing;

(b) the categories of personal data concerned;

(c) the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;

(d) where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period;

(e) the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing;

(f) the right to lodge a complaint with a supervisory authority;

(g) where the personal data are not collected from the data subject, any available information as to their source;

(h) the existence of automated decision-making, including profiling,”

Paragraph 3 states: *The controller shall provide a copy of the personal data undergoing processing. For any further copies requested by the data subject, the controller may charge a reasonable fee based on administrative costs. Where the data subject makes the request by electronic means, and unless otherwise*

requested by the data subject, the information shall be provided in a commonly used electronic form.

Section 3 refers to “Rectification and erasure”

Article 16 “Right to rectification” states: *The data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her. Taking into account the purposes of the processing, the data subject shall have the right to have incomplete personal data completed, including by means of providing a supplementary statement.*

Article 17 “Right to erasure” (‘right to be forgotten’) states: *The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:*

(a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;

(b) the data subject withdraws consent on which the processing is based.....

Section 4 refers to the “Right to object and automated individual decision-making”.

Article 21 “Right to object”, at paragraph No. 6, states: *Where personal data are processed for scientific or historical research purposes or statistical purposes pursuant to Article 89(1), the data subject, on grounds relating to his or her particular situation, shall have the right to object to processing of personal data concerning him or her, unless the processing is necessary for the performance of a task carried out for reasons of public interest.*

Chapter IV is dedicated to “Controller and processor”.

Section 1 “General obligations”. Article 24 refers to the “Responsibility of the controller”.

Section 2 is dedicated to “Security of personal data”.

Article 32, Security of processing

Paragraph 1 states: *Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:*

- (a) the pseudonymisation and encryption of personal data;*
- (b) the ability to ensure the on-going confidentiality, integrity, availability and resilience of processing systems and services;*
- (c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;*
- (d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.*

Paragraph 2 states: *In assessing the appropriate level of security account shall be taken in particular of the risks that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.*

Chapter VI “Independent supervisory authorities”

Section 1 “Independent status”, Article 51 refers to the “Supervisory authority”

Paragraph 1 states: *Each Member State shall provide for one or more independent public authorities to be responsible for monitoring the application of this Regulation, in order to protect the fundamental rights and freedoms of natural persons in relation to processing and to facilitate the free flow of personal data within the Union (‘supervisory authority’).*

Chapter IX “Provisions relating to specific processing situations”

Article 89 refers to “Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes”

Point 1 states: *Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in 4.5.2016 L 119/84 Official Journal of the European Union EN order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.*

Point 2 states: *Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far*

as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.

Point 3 states: Where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18, 19, 20 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.

Point 4 states: *Where processing referred to in paragraphs 2 and 3 serves at the same time another purpose, the derogations shall apply only to processing for the purposes referred to in those paragraphs.*

9. ECONOMIC CONSIDERATIONS

Cost considerations are essential before implementing a population-based registry. Without a valid surveillance system, it is not possible to plan and evaluate population health services, implement prevention interventions and identify “vulnerable” subgroups in terms of disease burden, such as the elderly, the young, the poor, the unemployed. Surveillance and evaluation involve a systematic way of learning from experience: these lessons learned should then be used to improve current activities and promote better planning by carefully selecting alternatives for future actions and resource allocation. The economic benefit of a good surveillance system clearly exceeds the cost of registers.

A population-based register may be costly. To produce meaningful data, it must remain in operation for at least one year, although a longer period is preferable and a continued operation is advisable. Therefore, register implementation is not possible if funds are not available.

Population-based registries that perform event data collection are more expensive, especially if hot pursuit is used. Besides the above mentioned costs, these types of registries also need funding for a detailed prospective data collection and for the validation of diagnostic information. Data collection includes: identification of patients, reading of medical records, making inquiries to additional data sources, data filing and validation. This means that a team of epidemiologists, nurses, medical doctors and informatics must be dedicated full time to this work. It should be recognised that this type of registers usually collect information that permits analyses of research questions beyond the monitoring of incidence, mortality and case fatality. This may concern the role of risk factors for disease prediction or the role of treatment for patient survival. An epidemiologic research team, the analysis of risk factors by linkage to HIS/HES data and treatment effects by linking

the registry to other data sources (e.g. data on drugs and invasive procedures) should be envisaged and included in the costs.

Registries based on record linkage between administrative databases are cost-effective, but they depend on the data quality of the hospital discharge and cause of death registers, as well as on the possibility of a valid record linkage. Moreover, these methods need further evaluation and implementation. Notably, if the hospital discharge and mortality registers are available for record linkage, costs for linkage and dissemination of results are low. The main costs for using this methodology to assess disease occurrence in a defined population concern the need to perform regular validations of the diagnostic information.

Sometimes access to data files (mortality, hospital discharges records), are not free of charges, so they produce additional costs.

10. CONCLUSIONS

Population based registries are intended for **researchers, health professionals and policy makers.**

The **objectives** of population based registries are: to evaluate the frequency and distribution of a disease in the population and provide indicators, such as incidence and survival rates and case-fatality; to evaluate trends and changing patterns; to monitor prevention programmes these registries provide a framework for research into aetiology. These registries may be important also to evaluate outcomes and treatment effectiveness.

Being focused on general population, a population based registry provides a comprehensive picture of a disease in the community, highlights problems, and suggests where treatment facilities are most in need of improvement.

Event identification and validation procedures depend on data collection methods (hot and cold pursuit), the healthcare system, the financing system of health care providers and the diagnostic criteria applied in the definition and validation of events. This process implies a great effort in training personnel, implementing quality control to read and collect information, classify events according to standardised diagnostic criteria, organise local site visits to assure that standard levels are respected and maintained.

The accuracy of rates is related to the completeness and quality control of the data collected per each numerator (all events from death certificates, hospital discharge registry) and denominator (census or population under surveillance). Completeness also depends on the possibility to trace the subjects treated outside the hospital (nursing homes, clinics, GPs) or outside the area of surveillance.

Results from population-based registries are available usually with a delay of 3-5 years, unlike current administrative data and statistics. Event identification and validation and the assessment of the vital status at different survival times make population-based registries very expensive and time consuming, therefore they can be usually maintained only for a limited period, in a defined population of a reasonable size, and to provide answers to the specific questions for which they were established.

The value of a registry must be examined at intervals to ensure that the objectives are still relevant and are met. If they are not, the objectives should be revised or the registry closed. The critical question is: “can this be done in any other way?” if the answer is yes, then the registry is probably a luxury.

11. THE NETWORK OF EXPERTS

Experts dealing with the different aspects of setting up and implementing population based registries have participated in the preparation of this manual: epidemiologists, statisticians, general practitioners, public health professionals and experts in ethical issues. Their expertise has covered various topics, such as case definition, record linkage of different sources of information, validation procedures, and assessment of ECHI indicators. The goal was to promote a fruitful discussion for the preparation of this operative manual and these guidelines to train the personnel involved in population based registries.

In order to encourage intense long-distance exchange, a web-based virtual platform, called the WP8 Community of Practice (CoP), was created so that these experts could interact and communicate; the CoP encouraged the transfer of knowledge, the development of new ideas, the re-framing of problems and the finding of original solutions.

(<http://wp8community.bridgehealth.eu/login/index.php>).

The CoP was based on an open-source software, developed through the MOODLE learning platform, through which propositions were developed and the decision-making process fostered. The Web Platform was an online system with several services tailored on the WP8-task1 needs, and an open forum to discuss proposals about each issue was set up. The activity necessary to produce the expected deliverables was enriched by the discussion carried out in each forum and in meetings organised with the support of the Ministry of Health [see the Report: *"Creation and development of the Italian network supporting the European BRIDGE-Health project aimed at structuring and providing sustainability to European activities in the field of Health Information (HI)"* www.cuore.iss.it].

This product is based on the work of the experts; they performed a literature review, collected information and participated as writing group:

Name	Affiliation
Sandro Bonfigli	Ministry of Health
Ovidio Brignoli	Italian College of General Practitioners
Alessandra Burgio	National Institute of Statistics
Flavia Carle	Ministry of Health
Antonio Carraturo	Cancer Registry of Latina
Susanna Conti	Istituto Superiore di Sanità
Giovanni Corrao	University of Milano Bicocca
Roberta De Angelis	Istituto Superiore di Sanità
Lidia Di Minco	Ministry of Health
Serena Donati	Istituto Superiore di Sanità
Marco Ferrario	University of Insubria, Varese
Daniela Galeone	Ministry of Health
Lidia Gargiulo	National Institute of Statistics
Silvia Ghirini	Istituto Superiore di Sanità
Simona Giampaoli	Istituto Superiore di Sanità
Laura Iannucci	National Institute of Statistics
Ilaria Luzzi	Istituto Superiore di Sanità
Marina Maggini	Istituto Superiore di Sanità
Aldo Pietro Maggioni	European Society of Cardiology
Antonella Olivieri	Istituto Superiore di Sanità
Monica Pace	National Institute of Statistics
Luigi Palmieri	Istituto Superiore di Sanità
Fabio PannoZZo	Cancer Registry of Latina
Alessio Pitidis	Istituto Superiore di Sanità
Andrea Ricci	Istituto Superiore di Sanità
Emanuele Scafato	Istituto Superiore di Sanità
Alessandro Solipaca	National Institute of Statistics
Lorenzo Spizzichino	Ministry of Health
Antonia Stazi	Istituto Superiore di Sanità
Domenica Taruscio	Istituto Superiore di Sanità
Virgilia Toccaceli	Istituto Superiore di Sanità
Marina Torre	Istituto Superiore di Sanità
Giuseppe Traversa	Istituto Superiore di Sanità
Luigi Uccioli	Tor Vergata University, Rome
Diego Vanuzzo	ANMCO, Heart Care Foundation
Giovanni Veronesi	University of Insubria, Varese
Luciano Vittozzi	Istituto Superiore di Sanità

12. REFERENCES

1. Health at a glance: Europe 2016 state of health in the EU cycle. Paris: OECD Publishing, 2016.
2. http://ec.europa.eu/health/indicators/indicators_en
3. Public Health Surveillance WHO at www.who.int/topics/public_health_surveillance
4. <http://www.bridge-health.eu>
5. Agency for Healthcare Research and Quality. Effective Health Care Programme: Registries for evaluating patient outcomes: a User's guide. 3rd Edition, 2014. www.effectivehealthcare.ahrq.gov/registries-guide-3.cfm
6. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Geneva: WHO, 2013.
7. World Health Organization. Global status report on noncommunicable diseases 2010. Geneva: WHO, 2010.
8. IDC-Health Insight. White paper, September 2014; #IDCWP26W.
9. Kesteloot H, Sans S, and Kromhout D. Dynamics of cardiovascular and all-cause mortality in western and Eastern Union between 1970 and 2000. *Eur Heart J* 2006; 27: 107-113.
10. Global Health Observatory, www.who.int
11. Strong K, Mathers C, Leeder S, Beaglehole R. Preventing chronic diseases: how many lives can we save? *Lancet* 2005; 366: 1578-1582.
12. European Commission and Economic Policy Committee. The 2015 ageing report: underlying assumption and projection methodologies. 2014 European Economy No.8/2014.
13. Stamler L, Low Risk - and the "No More Than 50%" Myth/Dogma. *Arch Intern Med*, 2007, 167:537-539
14. http://www.healthindicators.eu/healthindicators/object_document/o5873n28314.htm
15. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to

the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

16. Tunstall-Pedoe H. Diagnosis, measurements and surveillance of coronary events. *Int J Epidemiol* 1989; 18 (Suppl 1): 169-173.
17. Tunstall-Pedoe H. Problems with criteria and quality control in the registration of coronary events in the MONICA Study. *Acta Med Scand Suppl* 1998; 728: 17-25.
18. Parkin D. The evolution of the population-based cancer registry. *Nat Rev* 2006; 6: 603-612
19. D'Elia L, Galletti F, Strazzullo P. Dietary salt intake and risk of gastric cancer. *Cancer Treat Res*. 2014;159:83-95
20. Cummings H, Birgham A. Diet and the prevention of cancer. *Brit Med J* 1998; 317: 1636-1640
21. PARENT. Methodological guidelines and recommendations for efficient and rational governance of patient registries. 2015 www.patientregistries.eu
22. World Health Organisation Regional Office for Europe. Myocardial Infarction Community Registers. Copenhagen: WHO, 1976, Public Health in Europe report No 5.
23. 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: WHO, 2003.
24. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M, Tolonen H, Ruokokoski E, Amouyel P. Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. Monitoring trends and determinants in cardiovascular disease. *Lancet* 1999; 353: 1547-1557.
25. Beaglehole R, Stewart A, Butler M. Comparability of old and new world health organization criteria for definite myocardial infarction. *Int J Epidemiol* 1987, 16 N.3, 373-376
26. The Joint European Society of Cardiology/American College of Cardiology Committee. Myocardial infarction redefined. A consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the Redefinition of Myocardial Infarction. *Eur Heart J* 2000;
27. Luepker VR, Apple FS, Chistenson RH, Crow RS, Fortmann SP, Goff D, Goldberg RJ, Hand MM, Jaffe AS, Julian DG, Levy D, Manolio T, Mendis S, Mensah G, Pajak A, Prineas R, Reddy S, Roger V, Rosamond WO, Shahar E,

- Sharrett R, Sorlie P, Tunstall-Pedoe H. Case definitions for acute coronary heart disease in epidemiology and clinical research studies. *Circulation* 2003; 108: 2543-2549.
28. Fox KA, Birkhead J, Wilcox R, Knight C, Barth J. British Cardiac Society Working Group on the definition of myocardial infarction. *Heart* 2004; 90: 603-609.
29. The EUROCISS Working Group. Coronary and cerebrovascular population-based registers in Europe: are morbidity indicators comparable? Results from the EUROCISS Project. *Eur J Public Health* 2003; 13 (Suppl 3): 55-60.
30. European Community Health Indicators Monitoring - ECHIM Project. ECHI indicator development and documentation. Joint Action Development. Final report. 2012 (<http://www.echim.org>)
31. Madsen M, Gudnason V, Pajak A, Palmieri L, Rocha EC, Salomaa V, Sans S, Steinbach K, Vanuzzo D, EUROCISS Research Group. Population-based register of acute myocardial infarction: manual of operations. *Eur J Cardiovasc Prev Rehabil* 2007; 14 (Suppl 3): S3-S22.
32. Giampaoli S, Hammar N, Adany R, De Peretti C, EUROCISS Research Group. Population-based register of stroke: manual of operations. *Eur J Cardiovasc Prev Rehabil* 2007; 14: S23-S41.
33. Primatesta P, Allender S, Ciccarelli P, Doring A, Graff-Iversen S, Holub J, Panico S, Trichopoulos A, Verschuren WM; EUROCISS Research Group. Cardiovascular surveys: manual of operations. *Eur J Cardiovasc Prev Rehabil*. 2007 Dec;14 Suppl 3:S43-61
34. 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. *Int J Epidemiol* 2001; 30 (Suppl 1): S23-S29
35. Gruppo di Ricerca del Progetto Registro per gli Eventi Coronarici e Cerebrovascolari. Registro Nazionale italiano degli eventi coronarici maggiori: tassi di attacco e letalità nelle diverse aree del paese. *Ital Heart J* 2005; 6: 667-673.
36. Palmieri L, Barchielli A, Cesana G, de Campora E, Goldoni CA, Spolaore P, Uguccioni M, Vancheri F, Vanuzzo D, Ciccarelli P, Giampaoli S. The Italian Register of Cardiovascular Diseases: attack rates and case fatality for cerebrovascular events. *Cerebrovasc Dis* 2007, 24: 530-539.
37. Wagner G. The role of the registry in cancer control. Eds Parkin DM, Wagner G and Muir CS n. 66 IARC Lyon 1085.

38. International Association of Cancer Registries. A history. 2010. IARC Supplement, vol 10: 1-50.
39. International Classification of Diseases for Oncology 2000. Geneva: WHO, 2000.
40. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, Trama A, Visser O, Brenner H, Ardanaz E, Bielska-Lasota M, Engholm G, Nennecke A, Siesling S, Berrino F, Capocaccia R; EUROCARE-5 Working Group. are of little value if they are accompanied by disability and pain. Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE--5-a population-based study. *Lancet Oncol.* 2014 Jan;15(1):23-34.
41. <http://www.eurordis.org/about-rare-diseases>
42. www.orpha.net
43. Mazzucato M, Visonà Dalla Pozza L, Manea S, Minichiello C, Facchin P. A population-based registry as a source of health indicators for rare diseases: the ten-year experience of the Veneto Region's rare diseases registry. *Orphanet J Rare Dis* 2014; 9: 37.
44. Holder Y, Peden M, Krug E et al (Eds). Injury surveillance guidelines. Geneva, World Health Organization, 2001. http://www.who.int/violence_injury_prevention/publications/surveillance/surveillance_guidelines/en/
45. WHO Injury Surveillance Guidelines for Less-Resourced Environments, 2001, CDC <https://www.cdc.gov/nchs/ppt/ice/mdataset-.ppt>
46. https://ec.europa.eu/health/data_collection/databases/idb_en
47. EURODIAB ACE Study Group. Variation and trends in incidence of childhood diabetes in Europe. *Lancet* 2000; 355: 873-876.
48. Patterson C, Dalhquist GG, Gyurus E, Green A, Soltész G; EURODIAB Study Group. Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet* 2009; 373: 2027-2033.
49. The DIAMOND Project Group. Incidence and trends of childhood type 1 diabetes worldwide 1990-1999. *Diabet. Med.* 2006; 23, 857-866.
50. Bruno G, La Porte R, Merletti F, Biggeri A, McCarty D, Pagano G. National diabetes programmes. Application of capture-recapture to count diabetes? *Diabetes Care* 1994; 17: 548-556.

51. Carle F, Gesuita R, Bruno G, Coppa GV, Falorni A, Lorini R, Martinucci ME, Pozzilli P, Prisco F, Songini M, Tenconi MT, Cherubini V; RIDI Study Group. Diabetes incidence in 0- to 14-year age-group in Italy: a 10-year prospective study. *Diabetes Care* 2004; 27: 2790-2796.
52. Bruno G, Maule M, Merletti F, Novelli G, Falorni A, Iannilli A, Iughetti L, Altobelli E, d'Annunzio G, Piffer S, Pozzilli P, Iafusco D, Songini M, Roncarolo F, Toni S, Carle F, Cherubini V; RIDI Study Group. Age-period-cohort analysis of 1990-2003 incidence time trends of childhood diabetes in Italy: the RIDI study. *Diabetes* 2010; 59: 2281-2287.
53. 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, Massi Benedetti M. Core Standards of the EUBIROD Project. Defining a European diabetes data dictionary for clinical audit and healthcare delivery. *Methods Inf Med* 2016; 55:166-175.
54. <http://www.iss.it/rnic/>
55. Olivieri A; Italian Study Group for Congenital Hypothyroidism. The Italian National Register of infants with congenital hypothyroidism: twenty years of surveillance and study of congenital hypothyroidism. *Ital J Pediatr* 2009; 35: 2.
56. Olivieri A; Italian Study Group for Congenital Hypothyroidism. Epidemiology of congenital hypothyroidism: what can be deduced from the Italian registry of infants with congenital hypothyroidism. *J Matern Fetal Neonatal Med* 2012; 25 (Suppl 5): 7-9.
57. www.iss.it/gemelli
58. Stazi MA, Cotichini R, Patriarca V, Brescianini S, Fagnani C, D'Ippolito C, Cannoni S, Ristori G, Salvetti M. The Italian Twin Project: from the personal identification number to a national twin registry. *Twin Res* 2002; 5: 382-386.
59. Donati S, Maraschini A, Buoncristiano M, and the Regional maternal mortality working group. Methods to estimate maternal mortality: a global perspective. *J Epidemiol Community Health* 2015. pii: jech-2015-205547. doi:10.1136/jech-2015-205547.
60. Donati S, Maraschini A, Lega I, Basevi V, Buoncristiano M. Do generic correction algorithms produce reliable estimates? *Lancet* 2016;387:1815-6.
61. Donati S. Maternal mortality estimates. *Lancet* 2014;384(9961):2210. doi: 10.1016/S0140-6736 (14)62422-3.

62. Donati S, Senatore S, Ronconi A, and the Regional maternal mortality working group. Maternal Mortality in Italy: a record linkage study. *BJOG* 2011; 118:872-9.
63. Donati S, Maraschini A, Lega I et al. Convegno Sorveglianza della mortalità e grave morbosità materna in Italia - Istituto Superiore di Sanità - Roma, 27 maggio 2016. *Not Ist Super Sanità* 2016;29;11-5
64. Leonardsson O, Garellick G, Karrholm J, Akersson K, Rogmark C. Changes in implant choice and surgical technique for hemiarthroplasty: 21,436 procedures from the Swedish Hip Arthroplasty Register 2005-2009. *Acta Orthop* 2012; 83: 7-13.
65. RIAP Progetto Registro Italiano ArtroProtesi. Idea, Sviluppo e avvio Roma: Il Pensiero Scientifico Editore, 2014.
66. Torre M, Romanini E, Zanoli G, Carrani E, Luzi I, Leone L, Bellino S. Monitoring Outcome of Joint Arthroplasty in Italy: Implementation of the National Registry. *Joints* 2017;5:70-78
67. Nanni Costa A, De Cillia C, Di Ciaccio P, Rizzato L, Venettoni S. Solid Organ Transplantation: the Italian Experience in the International Context. *Transplantation Proceedings*. 2011, 43:2446-2447
68. <http://trapianti.net/en/>
69. <http://patientregistries.eu/lor>
70. WHO STEPS Stroke Manual: the WHO STEP wise approach to stroke surveillance. <http://www.who.int/chp/steps/Manual.pdf>
71. Federici A. Le parole della nuova Sanità. Roma: Il Pensiero Scientifico Editore, 2002.
72. Corrao G, Mancina G. Generating evidence from computerized healthcare utilization database. *Hypertension* 2015; 65: 490-498.
73. Mahonen M, Salomaa V, Keskimaki I, Moltchanov V. The feasibility of routine mortality and morbidity register data linkage to study the occurrence of acute coronary heart disease events in Finland. The Finnish Cardiovascular Diseases Registers (CVDR) Project. *Eur J Epidemiol* 2000; 16: 701-711.
74. Madsen M, Davidsen M, Rasmussen S, Abildstrøm SZ, Osler M. The validity of the diagnosis acute myocardial infarction in routine statistics. A comparison of mortality data and hospital discharge data with the Danish MONICA registration. *J Clin Epidemiol* 2003; 56: 124-130.

75. Hammar N, Alfredsson L, Rosén M, Spetz CL, Kahan T, Ysberg AS. A national record linkage to study acute myocardial infarction incidence and case fatality in Sweden. *Int J Epidemiol* 2001; 30 (Suppl 1): S30-S34.
76. Pajunen P, Koukkunen H, Ketonen M, Jerkkola T, Immonen-Raiha P, Kärjä-Koskenkari P, Mähönen M, Niemelä M, Kuulasmaa K, Palomäki P, Mustonen J, Lehtonen A, Arstila M, Vuorenmaa T, Lehto S, Miettinen H, Torppa J, Tuomilehto J, Kesäniemi YA, Pyörälä K, Salomaa V. The validity of the Finnish Hospital Discharge Register and the Causes of Death Register data on coronary heart disease. *Eur J Cardiovasc Prev Rehabil* 2005; 12 (Suppl 2): 132-137.
77. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN; Guidelines Working Group for Disease Control and Prevention (CDC). Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm Rep* 2001;50(RR-13):1-35; quiz CE1-7.
78. Project Management Institute. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*. 4th ed. Newtown Square, PA: Project Management Institute; 2008.
79. Dreyer NA, Garner S. Registries for robust evidence. *JAMA* 2009; 302: 790-791.
80. Solomon DJ, Henry RC, Hogan JG, Van Amburg GH, Taylor J. Evaluation and implementation of public health registries. *Public Health Rep* 1991; 106: 142-150.
81. Glaser SL, Clarke CA, Gomez SL, O'Malley CD, Purdie DM, West DW. Cancer surveillance research: a vital subdiscipline of cancer epidemiology. *Cancer Causes Control* 2005; 16: 1009-1019.
82. Bookman MA. Using tumor registry resources in analyzing concordance with guidelines and outcomes. *Oncology (Williston Park)* 2000; 14 (11A): 104-107.
83. Picciotto S, Forastiere F, Stafoggia M, D'Ippoliti D, Ancona C, Perucci CA. Associations of area based deprivation status and individual educational attainment with incidence, treatment, and prognosis of first coronary event in Rome, Italy. *J Epidemiol Community Health* 2006;60:37-43
84. Lyratzopoulos G, Patrick H, Campbell B. Registers needed for new interventional procedures. *Lancet* 2008; 371: 1734-1736.
85. Fellegi P, Sunter B. A theory for record linkage. *J Am Stat Ass*, 1969, 64: 1183-1210

86. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication. <http://www.icmje.org>. Accessed August 14, 2012.
87. Wittes JT, Colton T, Sidel VW. Capture-recapture methods for assessing the completeness of ascertainment when using multiple information sources. *J Chronic Dis* 1974. 27: 25-36
88. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *Eur J Cancer* 2009; 45: 756-764.
89. Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. AHRQ Publication No. 13(14)-EHC111 April 2014 Volume 1 Section II Legal and Ethical Considerations for Registries