Innovative use of data sources: A Cross-sectional study of Data Linkage Practices across European Countries

CURRENT STATUS: UNDER REVIEW

Romana Haneef
Sante publique France
romana.haneef@gmail.com Corresponding Author
ORCiD: https://orcid.org/0000-0001-7741-0268

Marie Delnord
Sciensano

Michel Vernay
Sante publique France

Emmanuelle Bauchet
Sante publique France

Rita Gaidelyte
National Institute of Hygiene and Health information centre

Herman Van Oyen
Sciensano

Zeynep Or
Institute of Research and Information for Health Economics

Beatriz Pérez-Gómez
Instituto de Salud Carlos III

Luigi Palmieri
National Institute of Health

Peter Achterberg
National Institute of Public Health and the Environment (RIVM)

Mariken Tijhuis
National Institute of Public Health and the Environment (RIVM)
Metka Zaletel  
Institut national de sante publique du Quebec

Stefan Mathis-Edenhofer  
The Austrian National Public Health Institute

Ondřej Májek  
National Research Council Canada

Håkon Haaheim  
The Norwegian Directorate of Health

Hanna Tolonen  
Finnish Institute of Health and Walfare (THL)

Anne Gallay  
Sante Publique France

DOI:  
10.21203/rs.2.22164/v1

SUBJECT AREAS  
Health Economics & Outcomes Research  
Artificial Intelligence and Machine Learning

KEYWORDS  
Innovation, linked data, artificial intelligence, machine learning technique, health status monitoring, public health surveillance, health information, health indicators
Abstract
Background The availability of data generated from different sources is increasing with the possibility to link these data sources together. However, linked administrative data can be complex to use and may require advanced expertise and skills in statistical analysis. The main objectives of this study were to describe the current use of data linkage at the individual level and the artificial intelligence (AI) in routine public health activities, and to identify the related health outcome and intervention indicators and determinants of health for non-communicable diseases.

Method We performed a survey across European countries to explore the current practices applied by national institutes of public health and health information and statistics for innovative use of data sources (i.e., the use of data linkage and/or the AI).

Results The use of data linkage and the AI at national institutes of public health and health information and statistics in Europe varies. The majority of European countries use data linkage in routine by applying a deterministic method or a combination of two types of linkages (i.e., deterministic & probabilistic) for public health surveillance and research purposes. The use of AI to estimate health indicators is not frequent at national institutes of public health and health information and statistics. Using linked data, 46 health outcome indicators related to seven health conditions, 34 indicators related to determinants and 23 to health interventions were estimated in routine. Complex data regulation laws, lack of human resources, skills and problems with data governance, were reported by European countries as obstacles to link different data sources in routine for public health surveillance and research.

Conclusions Our results highlight that the majority of European countries have integrated data linkage in routine public health activities but a few use the AI. A sustainable national health information system and a robust data governance framework allowing to link different data sources are essential to support evidence-informed health policy development process. Building analytical capacity and awareness of the added value of data linkage in national institutes is necessary for improving the utilization of linked data in order to improve the monitoring of public health activities.

Background
The availability of administrative data generated from different sources is increasing and the possibility to link these data sources with other databases offers unique opportunities to answer those research questions which require a large sample size or detailed data on hard-to-reach populations [1]. Data linkage can generate evidence with a high level of external validity and applicability for policymaking [1]. Over an extensive period, these population data (i.e., linked administrative data) can ensure high statistical power, reducing methodological issues relating to attrition, recall bias and lost-of-follow up [2], allowing more detailed stratified analyses of subgroups according to age, or specific geographical regions, and provide rapid access to data collected in a standardized format [3-5].

The value of any surveillance system ultimately depends on timely and reliable information [6]. There are several data sources which are used for public health surveillance, for example, health interviews and examination surveys, diseases-specific registries, epidemiological cohort studies, hospital discharge data, health insurance claims, mortality database, etc. Traditional data sources (e.g., health interview and examination surveys, disease-specific registries, etc.) and administrative data sources (e.g., hospital discharge, health insurance claims, causes of mortality data, etc.) complement each other and can increase the completeness and comprehensiveness of health information by taking into account various dimensions of health and risk factors influencing health status directly and indirectly. Linking various data sources improves completeness and comprehensiveness of information to guide health policy process, effective patient care and health services management [7]. Data linkage is an important technique that connects detailed individual-level information from different data sources. This methodology potentiate the capacity to study disease burden and progression, risk factors, care pathways and long-term outcomes for public health research and health surveillance [1]. However, linked administrative data can be complex to use and may require advanced expertise and skills in statistical analysis [8]. Generating efficiently comparable and timely health information across European Union (EU), European Economic Area (EEA) and other European countries requires to perform data linkage and apply AI to estimate health indicators. Many countries have already invested in data linkage to improve their health information system [9], but there are wide differences
in capacity across European countries to perform data linkage in routine.

We explored the differential use of data linkage in routine health monitoring based on the latest developments in new methods and analysis across European countries. This study was carried out under the InfAct (Information for Action) [10] which is a joint action of Member States aiming to develop a more sustainable EU health information system through improving the availability of comparable, robust and policy-relevant health status data and health system performance information. InfAct gathers 40 national health authorities from 28 Member States (MSs). This study is part of a work package (WP9) focused on innovation in health information system (i.e., using data linkages and/or AI) to improve public health surveillance and health system performance for health policy development process. The main objectives of this study were 1. to describe the current use of data linkage at the individual level and the AI techniques applied in routine public health activities and 2. to identify the relevant health outcome and intervention indicators estimated and determinants of health for non-communicable diseases.

Methodology

We performed following steps to achieve the objectives of this study:

1. **Literature search**

We reviewed the existing literature published on the use of data linkage and the AI (i.e., one technique of AI is machine learning technique) for health status monitoring using PubMed on Dec. 1, 2018. We included in our search peer-reviewed articles, systematic reviews and published reports published in English language. The search strategies used are reported in additional file 1. Based on the review, we identified different data sources used for data linkage, the use of artificial intelligence [AI]), health outcome and intervention indicators and determinants of health (Additional file2). This was not an exhaustive search and was performed only to identify any existing questionnaire or relevant information to be used to develop a questionnaire to identify the current practices in innovative use of data sources across European countries.

2. **Definition of innovative use of data sources**

We developed the definition of “innovative use of data sources” in the context of public health and
health information system and defined as:

*The linkage of different data sources* (health surveys and/or disease-specific or population-based registries and/or national cohort and/or clinical research datasets and/or administrative data and/or electronic health records and/or X-data sources i.e., information on determinants of health and can include data on various exposures [Additional file 2]) with each other using linkage technology and/or *The use of AI* either to linked data or to an individual data set, allowing a better understanding of what determines population health or to promote the efficiency of the health system and guide decision making at different geographical levels or at other categorization parameter level.

3. **Development of web-based survey**

We developed a questionnaire and requested information on data sources used for linkage, general characteristics of data linkage, the use of AI to estimate health indicators, related health outcomes and intervention indicators estimated and determinants of health of non-communicable diseases *(Additional file 3)*. We reported these results according to the status of using data linkage or the AI in routine public health activities across European countries into following three categories: 1. **Advanced** (i.e., those who use data linkage or the AI in routine to estimate health indicators), 2. **In Progress** (i.e., those for whom the deployment of these innovative techniques [i.e., data linkage or the AI] is still underway and expect to integrate these techniques in next 5 years), and 3. **Not yet** (i.e., those for which the use of these techniques are not foreseen yet).

Survey participants were asked to report at least three health indicators related to priority medical conditions in their country. We adopted the Euro-REACH Framework (i.e., it is project based on an international collaboration to improve access to health care data through cross-country comparisons) [11] to classify the identified health outcome indicators, determinants of health and intervention indicators under following categories: health outcome indicators (1. Health characteristics, 2. Mortality, 3. Human function and quality of life and 4. Life expectancy and well-being), determinants of health (1. Physical environment, 2. Socioeconomic and environment, 3. Health behavior and lifestyle and 4. Biological /metabolic parameters) and intervention indicators (1. Prevention, 2. Promotion and 3. Others). We also asked specific information about objective of health indicators estimation (i.e., for public health monitoring, scientific research [clinical, epidemiology, public health], both),
status of their use (i.e., was used, currently in use or could be produced in future) and level of estimation (i.e., national, sub-national, metropolitan, at all levels). If the same health indicator was reported more than once either as being estimated currently or to be estimated in future by different countries, we counted those health indicators once. The web-based questionnaire was developed using the Lime Survey tool by Data lab of Santé Publique France. The questionnaire includes both closed and open-ended questions (i.e., 20 questions) *(Additional file 3)*. This questionnaire was reviewed by a group of experts in health information systems in their country and revised according to their feedback before the launch of the survey. The web-based version of the questionnaire was pretested by the co-authors (SME, RH and RG) from respective national public health institutes of Austria, France and Lithuania to check the visibility of the questions and contents.

The survey participants were the partners of the InfAct project and/or national representatives or experts or advisors for health information in their countries and could be either employed by government, national institutes of public health and health information and statistics or research departments of the universities.

The invitation email with an electronic link to the questionnaire was sent on April 1, 2019, to the identified representatives in 31 European countries to complete the survey in four weeks (i.e., April 30, 2019) *(Additional file 4)*. The first reminder through email was sent after one month of survey launch, on May 3, 2019, and the second reminder after two weeks of the first reminder, on May 23, 2019. The abbreviations of the member countries and the names of the survey respondents are reported in *additional file 5*.

**Study outcomes**

The main outcomes of this study were the current practices in data linkage and the AI and related health indicators estimated in routine public health activities across European countries. A descriptive analysis of the web-based questionnaire results has been performed using Microsoft Excel.

**Results**

1. **Literature search**

We reviewed 137 citations from PubMed and four reports from the following organizations: OECD
**Organization for Economic Co-operation and Development** [12], **Euro-REACH** [13], **HBM4EU (Human Biomonitoring for Europe)** [14], **EUROCISS (European Cardiovascular Indicators Surveillance Set)** [15], to develop this questionnaire (Fig. 1).

Fig. 1: Flow diagram of studies using linked data and artificial intelligence for health status monitoring to develop a questionnaire identifying various practices of data linkages across European countries in 2019 (insert here)

Thirty-one European countries [28 EU-MSs + 2 EEA (Iceland and Norway) + Others (Serbia)] were invited to participate in the WP9 survey and twenty-nine countries (i.e., EU MSs 27 + EEA 1 [Norway] + Others 1 [Serbia]) participated with a response rate of 94% (29/31). Hungary, Iceland and Northern Ireland did not participate. For the United Kingdom, data were provided separately by the three countries England, Scotland and Wales but was counted as one member state. The results have been validated by the all survey respondents.

2. **Use of data linkage in routine public health activities**

Our survey results highlighted that 24 European countries perform data linkage in their routine public health activities. These countries link administrative data such as EHRs, mortality data, disease specific registries whereas six of them (Cyprus, Italy, Poland, Portugal, Spain and Slovakia) are also developing this technique further to link with different other data sources (i.e., demographic data, domestic/leisure accidents data, congenital anomalies registry). Ireland and Latvia have ongoing initiatives of data linkage (*Table 1.1*).

*Table 1.1: Current status of European countries using data linkage in routine public health activities in 2019 for innovative use of data sources*

| Use of Data Linkage | Advanced N = 24 | In progress* N = 8 | Not yet N = 3 |
|---------------------|-----------------|--------------------|---------------|
| European Countries  | AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, HR, IT, LT, MT, NL, NO, PL, PT, SI, SK, SRB, SW, UK (ENG, SC, WL) | CY, ES, IE, IT, PL, PT, SK, LV | GR, LU, RO |

* 6 countries (CY, ES, IT, PL, PT & SK) use data linkage in routine (i.e., advanced) but also developing further this technology to link different other data sources (i.e., in progress).
Three countries (Greece, Luxembourg and Romania) have not yet planned any perspectives to integrate data linkage in routine public health activities. Following reasons were mentioned by some countries for not having institutionalized data linkage in their country: lack of a public health institution which should collect and govern the health related data, data linkage is not part of the health agenda, lack of commitment from the ministry of health, lack of resources to establish a national health information system, and the institutional complexity of the Ministry of Health and strict laws and regulations which hinder data linkage with different data sources.

**Objectives of data linkage:** Data linkage are performed in routine for different objectives such as for health status monitoring, health system performance, health policy or for scientific research (i.e., public health, epidemiology or clinical) purposes. Our results showed that data linkage was performed for health status monitoring in 20 countries (BE, CY, CZ, DE, DK, EE, ES, FI, FR, HR, IT, LT, MT, NL, PT, SI, SK, SRB, SW, UK (SC, WL), for health policy development in 13 (AT, BE, BG, DK, EE, FR, MT, NL, NO, PL, SK, SW, UK (SC, WL) and for scientific research (public health, epidemiological and clinical) purposes in 13 (BE, CZ, DE, DK, EE, ES, FI, FR, NL, PT, SI, SW, UK (ENG, SC, WL). Finland, Spain, Sweden and Scotland also perform data linkages to identify the risk factors. In Sweden, data linkage is also used to monitor compliance with national treatment guidelines to improve health care quality.

**Data sources used for linkage:** Our results showed that 24 European countries who perform data linkage in routine, used most frequently five following data sources: health-related administrative data sources, non-health related administrative data sources, disease-specific registries, national health surveys, population-based epidemiological cohort and clinical trials. *(Table 1.2).* These data sources are linked with each other in different combinations and some examples of various combinations used across member countries, are reported in *table 1.3.* These countries perform data linkage by using one of following information: social security number, patient unique identification number, person unique pseudonymous identifier, encrypted personal identification number, citizen or national identification number. In Ireland, the lack of a unique patient identifier number limits the potential to link with different data sources.

Table 1.2: Data sources used for linkage across European countries in 2019 for innovative use of data sources
| S/No | Data sources used for linkage | European countries |
|------|------------------------------|--------------------|
| 1    | **Health-related administrative data sources (i.e., Electronic Health Records)** | **Advanced N = 24** |
|      | Primary care visits, emergency care, referral records, hospital discharge, prescribed medications, health insurance claims, diagnostics procedures, laboratory tests, biobank | AT, BE, CY, DE, DK, EE, ES, FI, FR, HR, IT, LT, MT, NL, NO, PT, SI, SK, SRB, SW, UK[ENG, SC, WL] |
| 2    | **Non-health related administrative data sources ‡** | **In prog N = 2** |
|      | Birth and mortality database, education level, income tax, GIS, occupation, housing conditions, criminal statistics, land and housing, socioeconomic, census (demographic), house of handicap persons, environmental, road and transport, air pollution, UV light exposure | BE, CY, CZ, DE, DK, EE, ES, FI, FR, HR, IT, LT, MT, NL, NO, PL, PT, SI, SK, SRB, SW, UK[ENG, SC, WL] |
| 3    | **Disease-specific registries** | **LV** |
|      | Cancer, diabetes, cardiovascular, congenital malformation, tuberculosis, HIV/AIDS, inflammatory bowel disease, renal, reproductive health, dementia, organ transplantation, traffic accidents/trauma or injury, hospital registry of domestic and leisure accidents | BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, HR, IE, LV, MT, NL, NO, PL, PT, SK, SRB, SW, UK[ENG, SC, WL] |
| 4    | **National health surveys*** | **15** |
|      | National health examination and interview surveys | BE, CZ, DK, DE, EE, ES, FI, FR, IT, NL, NO, PT, SI, SW, UK[ENG, SC, WL] |
| 5    | **Population-based epidemiological cohort/National cohorts** | **7** |
|      | DANCOS, IDEFICS, CONSTANCE, ELFE, Growing up in Scotland, HealthWise Wales cohort, Millennium cohort, Caerphilly cohort study | DK, EE, FI, FR, NO, PL, UK[ENG, SC, WL] |
| 6    | **Clinical trials data** | **3** |
|      | FINGER, PRISOMATIC | DK, FI, UK[ENG, WL] |

* Latvia is developing data linkage techniques to link EHRs with other data sources.

‡ In Ireland, income database is linked with EHRs of prescribing medicine at small level. Latvia is developing data linkage techniques to link birth and mortality databases either with EHRs or with disease-specific registries.

* Poland is planning to link this national health survey data with other health data sources in near future. In Ireland, this is done for specific surveys such as housing and health conditions at small scale.

**General characteristics of linked dataset:** Our results showed that among 24 European countries who perform data linkage in routine, 17 do linkage at national level (Table 1.4). France, Portugal and Scotland do data linkage both at national and sub-national levels. Denmark, Germany, Norway and Sweden do data linkage at all levels. 23 countries either use the deterministic type of linkage (12 countries) or a combination of deterministic and probabilistic linkage (11 countries). Among 16/24
countries, linked data is available and is used in routine. Among 12/24 countries, the register owner (i.e., who governs the data register) provides the approval to access linked data. Among 15/24 countries, the accessibility to linked data is in routine or permanent whereas, in 13 countries, the accessibility could be ad-hoc or at intermittent basis depending on the project. Among 15/24 countries, linked data do not operate in real-time (i.e., integrate the updated information with minimum delay in time). Among 19/24 countries, linked data are flexible to integrate new variables. There are ongoing projects on data linkage (i.e., in next five years) aiming to integrate this technology in their routine public health activities in following European countries: Austria, Cyprus, Czech Republic, Ireland, Italy, Latvia, Norway, Poland, Portugal, and Spain.

Table 1.4: General characteristics of linked datasets in European countries in 2019 for innovative use of data sources (insert table here)

3. Use of artificial intelligence (AI) in routine public health activities

The use of AI is not frequent across European countries (Table 2). Only five countries have reported applying following techniques in routine public health activities: machine learning (Denmark, Finland, Sweden, and UK-Wales), natural language processing (Finland, Sweden, and UK-Wales), Markov decision process (Finland), support vector machine (Finland, UK-Wales), data mining (Finland) and TSP [Travelling Salesman Problem] modelling (Norway). Denmark can apply these techniques not only at a national level but also at a metropolitan level.

There are ongoing projects on the use of the AI (i.e., in next five years) to integrate this technology in routine public health activities in following countries: Croatia, Czech Republic, France, Germany, Norway, Portugal, and Spain. The objectives of these initiatives are for epidemiological research and surveillance of non-communicable and communicable disease estimating the prevalence and prediction of incidences of certain health conditions at various geographical levels.

Two countries mentioned that due to lack of human resources (Lithuania) and capacities/skills (Republic of Serbia) within their public health institutes, AI techniques are not applied in routine public health activities.

Some European countries also mentioned use of classical statistical techniques without the use of AI (Table 2).

Table 2: Current status of European countries using artificial intelligence in routine public health activities in 2019
4. Health indicators estimated using linked data

Using linked data, the majority of European countries estimate following health indicators:

**Health outcome indicators**

Participants were asked to select at least three health conditions and to report the related health outcome indicators which are most important for public health in their country. Using linked data, 46 health outcome indicators related to following seven health conditions were reported from 22 countries: cardiovascular (14), neurodegenerative disease (6), maternal and perinatal health (6), diabetes (6), suicide/trauma/injury (7), cancer (6) and hepatic failure (1) (*Table 3.1*). The main objectives to estimate these indicators were for public health monitoring and research purposes and the level of estimation was mainly at national and sub-national levels.

Table 3.1: Description of health outcome indicators estimated using linked data across European countries in 2019 (insert table here)

**Health determinants**

For the health determinants, participants were asked to report the corresponding determinants of the
identified health conditions. 34 health determinants related to various health conditions were reported by 15 member states (*Table 3.2*). These determinants are related to the physical environment (12), socioeconomic and environment (10), health behavior and lifestyle (6) and biological and metabolic parameters (3) (*Table 3.2*). These determinants were used to measure the potential associations between these risk factors and health conditions for public health monitoring and research purposes. These determinants can be stratified by age, sex, socioeconomic status and by area of residence.

Table 3.2: Description of health determinants using linked data across European countries in 2019 (insert table here)

*Health intervention indicators*

Participants were asked to report at least three health intervention indicators under three categories (i.e., prevention, promotion, others) corresponding to the given health conditions which are most important for public health in their country. Using linked data, 23 health intervention indicators related to following six health conditions were reported from 17 member states: maternal and perinatal health (7), cancer (6), diabetes (4), cardiovascular (2), neurodegenerative disease (2), suicide/trauma/injury (1) and lower/upper respiratory infections (1), (*Table 3.3*). The main objectives to estimate these indicators were to guide health policy process, public health monitoring and for research purposes. These intervention indicators are estimated mainly at national and sub-national levels and currently are in use.

Table 3.3: Description of health intervention indicators estimated using linked data across European countries in 2019 (insert table here)

**Discussion**

The results of this study showed variability in use of data linkage and the AI at national institutes of public health and health information and statistics across European countries. The majority of countries use data linkage in routine by applying either deterministic or a combination of two types of linkages (i.e., deterministic & probabilistic) for public health surveillance and research purposes. The use of a universal unique identifier, social security number or unique pseudonymous identifier is common to applying deterministic linkage technique among European countries. The use of AI is not frequent to estimate health indictors at national public health institutes. Across European countries, using data linkage, 46 health outcome indicators related to seven health conditions, 34 related
determinants and 23 health intervention indicators were reported. Some initiatives are ongoing as pilot projects to apply these techniques to improve health surveillance and to guide health policy development process.

A systematic review has shown some practices applied for data linkage in the field of perinatal health across Europe for health surveillance and research purposes [9]. Several other studies have explored various dynamics of population health such as social care, psychotic disorders, multi-morbidity, diabetes, obesity, mental health, cardiovascular, antibiotic use and Alzheimer using data linkage with different types of administrative data sources (both related to health and non-health) [7, 17–29]. For the surveillance of cancer, data linkage not only provides the opportunity to improve the population-based screening [30] but also helps in detecting different types of cancer recurrence [31] and evaluation of the socio-economic status of patients with cancer (e.g., return to work) [32]. Linked data also allows evaluating the interventions at various levels of the population [33]. The diversity and the volume of health information have been increasing rapidly and push to discover new parameters to improve population health with innovative approaches. In that context, some initiatives have been launched at the national levels to create health data hub/platform to be used for research and to guide the policy development process [34, 35].

There are some studies available which have discussed the advantages of using AI in early detection and diagnosis of certain conditions, treatment, as well as outcome prediction and prognosis evaluation with high precision [36, 37] but their use in population health to estimate and predict health indicators remain limited [38].

Our results highlighted that a few member countries have achieved the most advanced level in data linkage by linking health information (i.e., clinical, biobanks/laboratory tests, genetics) with education, occupation, housing quality, air pollution, criminal statistics and transport/road accidents, etc. This offers the exceptional opportunities to enrich information and to perform epidemiological research, health surveillance and consequently, guiding health policies to improve population health. However, majority of European countries have not reached that level in data linkage and the use of AI underlined following four main obstacles associated with the implementation and the use of data
linkage and advanced statistics: 1. The complex laws and data protection regulations which block linkage between different data sources with a deterministic approach (legal), 2. Lack of human resources and capacities/skills within national institutes of public health and health information statistics (technical), 3. Lack of governance of health information (data governance) and 4. Limited resources to support the health information infrastructure (organization and structural).

To address these gaps, we propose the following recommendations: A. Legal aspects: 1. more flexible data governance frameworks to support data linkage of different data sources should be encouraged [39], 2. Specific mandates to ensure data availability/access/capture and safe storage should be an integral part of the health information system, 3. Differences in implementation and the interpretation of the EU-GDPR (General Data Protection Regulations) and additional national regulations should be mapped and if possible harmonized across EU-MSs [40]; B. Technical aspects: 4. more collaborations and partnerships should be encouraged to build up capacities for use of new technology, to share new methods, skills, experiences and data for comparative research studies among EU national institutes of public health and health information statistics; C. Governance, 5. Initiatives to strengthen national health information infrastructure should be encouraged ; D. Organizational and structural aspects, 6. ministries of health and research in a member country should provide their support to develop national health data hubs/data platform to strengthen the national health information infrastructure.

There are a few limitations in this study. First, current practices of data linkages at national institutes of public health and health information and statistics we surveyed may differ from other research institutes in that country which we did not cover and might influence the results of this study. However, this survey provides the latest overview of current practices in data linkage and highlights the related obstacles in performing data linkage. Second, we limited the response burden of health indicators to three priority health conditions. Therefore, our results do not constitute an exhaustive list of health indicators that are used in the country to inform policy and practice. It may limit the number of health indicators being estimated using linked data and advanced statistics.

Conclusions
To our knowledge, this is the first study which provides the information about the current practices of
data linkage and the AI at national institutes of public health and health information and statistics across European countries. Our results highlight that the majority of the countries have integrated data linkage in routine public health activities but few use the AI. The European countries who are advanced in using both techniques data linkage and the AI could guide others by an exchange of their experiences and examples of good practices. A sustainable national health information system and data governance framework to link different data sources are essential to support evidence-based practices for health policy development process. Building analytical capacity in national institutes of public health and health information and statistics is necessary for improving the utilization of linked data in order to improve the monitoring of public health activities. These results ultimately contribute to strengthen the national health information system and would facilitate moving towards establishing an integrated EU- Health Information System.

List Of Abbreviations

AI: Artificial Intelligence
EU: European Union
EEA: European Economic Area
MSs: Member States
InfAct: Information for Action i.e., a joint action of Member States to establish a sustainable European health information system.
WP: Work Package
Euro-REACH: It is an international collaboration to improve access to health care data through cross-country comparisons.
OECD: Organization for Economic Co-operation and Development
HBM4EU: Human Biomonitoring for Europe
EUROCISS: European Cardiovascular Indicators Surveillance Set
PCA: Principal Component Analysis
GDPR: General Data Protection Regulations

Declarations
**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

All authors gave the consent for publication.

**Availability of data and materials**

Not applicable

**Competing interests**

All other authors declare that they have no competing interests related to the work.

**Funding**

This research has been carried out in the context of the project ‘801553 / InfAct’ which has received funding from the European Union’s Health Programme (2014-2020).

**Authors’ contributions**

Conceived and designed the survey: RH EB MV AG. Performed the study: RH AG. Analyzed the data: RH. Analysis of the data and interpretation of the results: RH AG. Contributed to the writing of the manuscript: All authors contributed to the writing of the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

We acknowledge the kind support of Jessica Gane (Data Department, Santé Publique France) for developing the web-based survey. We specially acknowledge Jean-Claude Desenclos and Yann Le Strat for their kind support in developing the survey questionnaire.

We acknowledge following partners who contributed to complete this survey: Bulgaria: Raina Nikolova (Public Health Institute), Croatia: Ivan Pristaš (CIPH, Cyprus: Vasos Scoutellas (Public health institute), Czech Republic: Šárka Daňková, Jiří Jarkovský (Institute of health information and statistics of Czech Republic), Denmark: Mette Bjerrum Koch (Danish Health data authority), Maja Bæksgaard Jørgensen (Statens Institute for Folkesundhed), Estonia: Eleri Lapp (MoSH), Finland (Mike Gissler (THL), Germany: Angelika Schaffrath Rosario (RKI), Greece: Spyridon Goulas (Governmental Health Insurance Organization), France: Jennifer Zeitlin (INSERM U1153), Sandrine Fosse (Santé Publique...
France (SPF), Laure Carcaillon-Bentata (SPF), Valérie Olie (SPF), Louis-Marie Paget (SPF), Viet Nguyen-Thanh (SPF), Ireland: Sheona Gilsenan (DoH), Italy: Brigid Unim (ISS), Latvia: Janis Misins (CDPC), Luxembourg: Anne-Charlotte Lorcy (Health Directorate), Malta: Neville Calleja (MFH), Norway: Heidi Lyshol (Norwegian Institute of Public Health), Inger Johanne Bakken (Norwegian Directorate of Health, Poland: Jakub Adamski (MoH), Piotr Nowosielski (MoH), Portugal: Paulo Nogueira (DGS), Martin José (Public Health Institute), Carlos Dias (Public health Institute), Romania: Silviu Radulescu (Public Health Institute), Serbia: Maja Krstic (Batut), Slovakia: Jan Cap (National Health Information Centre), Sweden: Hanne Lobosco (MoH), Rosita Claesson Wigand (Public Health Agency of Sweden), Jenny Börlin (Public Health Agency of Sweden), England: Robert Aldridge (Institute of Health Informatics, University of London), Scotland: Ian Grant (National Statistics Department) and Wales: Ronan Lyons (University of Swansea).

Disclaimer: This article presents independent research. The views expressed are those of the authors and not necessarily those of the Carlos III Institute of Health.

References
1. Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, Goldstein H: Challenges in administrative data linkage for research. Big Data Soc 2017, 4(2):2053951717745678-2053951717745678.
2. Ferrante A: The Use of Data-Linkage Methods in Criminal Justice Research: http://www5.austlii.edu.au/au/journals/CICrimJust/2009/3.html. Criminal Justice 2009.
3. GF R: Administrative and claims records as sources of health care cost data. Med Care 2009, 47(7 Suppl 1).
4. Charlton RA, Neville AJ, Jordan S, Pierini A, Damase-Michel C, Klungsøyr K, Andersen A-MN, Hansen AV, Gini R, Bos JHJ et al: Healthcare databases in Europe for studying medicine use and safety during pregnancy. Pharmacoepidemiology and Drug Safety 2014, 23(6):586-594.
5. Thygesen LC, Ersbøll AK: When the entire population is the sample: strengths and
limitations in register-based epidemiology. *European Journal of Epidemiology* 2014, 29(8):551-558.

6. WHO: Public Health Surveillance:
https://www.who.int/topics/public_health_surveillance/en/.

7. Lloyd K, McGregor J, John A, Craddock N, Walters JT, Linden D, Jones I, Bentall R, Lyons RA, Ford DV *et al*: A national population-based e-cohort of people with psychosis (PsyCymru) linking prospectively ascertained phenotypically rich and genetic data to routinely collected records: Overview, recruitment and linkage. *Schizophrenia Research* 2015, 166(1):131-136.

8. Bradley CJ, Penberthy L, Devers KJ, Holden DJ: Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future. *Health Services Research* 2010, 45(5p2):1468-1488.

9. Delnord M, Szamotulska K, Hindori-Mohangoo AD, Blondel B, Macfarlane AJ, Dattani N, Barona C, Berrut S, Zile I, Wood R *et al*: Linking databases on perinatal health: a review of the literature and current practices in Europe. *European Journal of Public Health* 2016, 26(3):422-430.

10. Joint Action on Health Information: https://www.inf-act.eu/. 2018.

11. EuroREACH: EuroREACH Framework: http://hdn.euhs-i.eu/performance/frameworks/euroreach-framework. 2013.

12. OECD: Health at Glance (OECD Indicators):
https://www.health.gov.il/PublicationsFiles/HealthataGlance2017.pdf. 2007.

13. Navigator HD: EuroREACH Framework: http://hdn.euhs-i.eu/performance/frameworks/euroreach-framework. 2013.

14. HBM4EU: Linking HBM, health surveys and registers:
https://www.hbm4eu.eu/deliverables/. 2018.
15. Eurociss: Cardiovascular Indicators Surveillance Set:
https://ec.europa.eu/health/ph_projects/2000/monitoring/fp_monitoring_2000_frep_10_en.pdf. 2000.

16. Health Research Board:
https://www.hrb.ie/fileadmin/publications_files/Proposals_for_an_Enableing_Data_Environment_for_Health_and_Related_Research_in_Ireland.pdf. 2019.

17. Lyons RA, Jones KH, John G, Brooks CJ, Verplancke J-P, Ford DV, Brown G, Leake K: The SAIL databank: linking multiple health and social care datasets. *BMC Medical Informatics and Decision Making* 2009, 9(1):3.

18. Tuppin P, Rudant J, Constantinou P, Gastaldi-Menager C, Rachas A, de Roquefeuil L, Maura G, Caillol H, Tajahmady A, Coste J et al: Value of a national administrative database to guide public decisions: From the systeme national d’information interregimes de l’Assurance Maladie (SNIIRAM) to the systeme national des donnees de sante (SNDS) in France. 2017(0398-7620 (Print)).

19. Chan Chee C, Chin F, Ha C, Beltzer N, Bonaldi C: Use of medical administrative data for the surveillance of psychotic disorders in France. *BMC Psychiatry* 2017, 17(1):386.

20. Rodgers Se Fau - Bailey R, Bailey R Fau - Johnson R, Johnson R Fau - Poortinga W, Poortinga W Fau - Smith R, Smith R Fau - Berridge D, Berridge D Fau - Anderson P, Anderson P Fau - Phillips C, Phillips C Fau - Lannon S, Lannon S Fau - Jones N, Jones N Fau - Dunstan FD et al: Health impact, and economic value, of meeting housing quality standards: a retrospective longitudinal data linkage study *Public Health Research* 2018.

21. Violán C, Foguet-Boreu Q, Hermosilla-Pérez E, Valderas JM, Bolíbar B, Fàbregas-Escarriola M, Brugulat-Guiteras P, Muñoz-Pérez MÁ: Comparison of the information
provided by electronic health records data and a population health survey to estimate prevalence of selected health conditions and multimorbidity. *BMC Public Health* 2013, 13(1):251.

22. Fuentes S, Cosson E, Mandereau-Bruno L, Fagot-Campagna A, Bernillon P, Goldberg M, Fosse-Edorh S, Group C-D: Identifying diabetes cases in health administrative databases: a validation study based on a large French cohort. *International Journal of Public Health* 2019, 64(3):441-450.

23. Orriols L, Delorme B, Gadegbeku B, Tricotel A, Contrand B, Laumon B, Salmi L-R, Lagarde E, on behalf of the Crg: Prescription Medicines and the Risk of Road Traffic Crashes: A French Registry-Based Study. *PLOS Medicine* 2010, 7(11):e1000366.

24. Mason KE, Pearce N, Cummins S: Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank. *The Lancet Public Health* 2018, 3(1):e24-e33.

25. Cleland B, Wallace J, Bond R, Black M, Mulvenna M, Rankin D, Tanney A: Insights into Antidepressant Prescribing Using Open Health Data. *Big Data Research* 2018, 12:41-48.

26. Gabet A, Danchin N, Puymirat E, Tuppin P, Olié V: Early and late case fatality after hospitalization for acute coronary syndrome in France, 2010-2015. *Archives of Cardiovascular Diseases* 2019, 112(12):754-764.

27. Williamson E DS, Morris S, Clarke CS, Thomas M, Evans H et al Risk of mortality and cardiovascular events following macrolide prescription in. *Rhinology* 2019, 57(4):252-260.

28. Hopkins C WE, Morris S, Clarke CS, Thomas M, Evans H, Little P et al Antibiotic usage in chronic rhinosinusitis: analysis of national primary care. *Rhinology* 2019, 6(10):136.
29. Ponjoan A G-OJ, Blanch J, Fages E, Alves-Cabreros L, et al: How well can electronic health records from primary care identify Alzheimer's. _Clin Epidemiol_ 2019, 11:509-518.

30. Mäjek O, Anttila A, Arbyn M, van Veen E-B, Engesæter B, Lönnberg S: The legal framework for European cervical cancer screening programmes. _European Journal of Public Health_ 2018, 29(2):345-350.

31. Hassett MJ, Uno H, Cronin AM, Carroll NM, Hornbrook MC, Ritzwoller D: Detecting Lung and Colorectal Cancer Recurrence Using Structured Clinical/Administrative Data to Enable Outcomes Research and Population Health Management. _Medical Care_ 2017, 55(12):e88-e98.

32. Kiasuwa-Mbengi RL, Nyaga V, Otter R, de Brouwer C, Bouland C: The EMPCAN study: protocol of a population-based cohort study on the evolution of the socio-economic position of workers with cancer. _Archives of Public Health_ 2019, 77(1):15.

33. Lyons RA, Turner S, Lyons J, Walters A, Snooks HA, Greenacre J, Humphreys C, Jones SJ: All Wales Injury Surveillance System revised: development of a population-based system to evaluate single-level and multilevel interventions. _Injury Prevention_ 2016, 22(Suppl 1):i50-i55.

34. Health FMo: Health Data Hub: https://www.health-data-hub.fr/?lang=en. 2019.

35. Government W: Welsh Government takes innovative approach to policy making in Wales: https://www.swansea.ac.uk/press-office/news-archive/2019/welshgovernmenttakesinnovativeapproachtopolicymakinginwales.php. 2019.

36. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, Abu-Hanna A: The coming of age of artificial intelligence in medicine. _Artificial Intelligence in Medicine_ 2009, 46(1):5-17.
37. Jha S, Topol EJ: Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. *JAMA* 2016, 316(22):2353-2354.

38. Flaxman AD, Vos T: Machine learning in population health: Opportunities and threats. *PLOS Medicine* 2018, 15(11):e1002702.

39. 2020 C: Data Protection and Artificial Intelligence: https://edps.europa.eu/data-protection/our-work/subjects/artificial-intelligence_en.

40. EPRS: How the General Data Protection Regulation changes the rules for scientific research:

   https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU(2019)634447_EN.pdf. 2019.

**Tables**

Table 1.3: Examples of different combinations of data linkages across European countries in 2019 for innovative use of data sources

| S/No | European countries          | Different combinations of data linkages (N = ~ 85)                                                                 |
|------|----------------------------|------------------------------------------------------------------------------------------------------------------------|
| 1    | Austria                    | Hospital discharge with outpatient visit (primary care visit)                                                            |
| 2    | Belgium                    | Hospital discharge with health insurance claim and educational attainment with mortality database                         |
|      |                            | Census with mortality database                                                                                        |
|      |                            | Health interview survey with mortality (cause-specific mortality, StatBEL)/prescribed medication (INAMI)              |
|      |                            | Disease-specific registries with mortality database                                                                    |
| 3    | Bulgaria                   | Registry of rare diseases with oncology registry                                                                         |
| 4    | Croatia                    | Primary health care visits with hospital discharge/health insurance claim/mortality (cause-specific mortality)        |
|      |                            | Cancer registry with geospatial registry                                                                               |
| 5    | Cyprus                     | Hospital discharge with mortality database                                                                            |
|      |                            | Cancer, diabetes, HIV/AIDS registries with mortality database                                                          |
| 6    | Czech Republic             | Health insurance claims with mortality database                                                                       |
|      |                            | Registry of hospitalizations is linked with mortality database                                                         |
|      |                            | Disease specific registries: Cancer, cardiovascular surgery and intervention, reproductive health registry
|      |                            | TBC registry, registry of injuries with mortality database/health insurance claims/registry of hospitalizations       |
| 7    | Denmark                    | National patient health register is linked with education, income, housing, transfer paym socioeconomic status, criminal statistics, etc.|
| 8    | Estonia                    | Hospital stay, primary and special ambulatory care linked with health insurance claims                                 |
|      |                            | Health insurance claims and prescriptions are linked with causes-specific mortality                                    |
|      |                            | Birth register linked with causes-specific mortality                                                                   |
|      |                            | Cancer, tuberculosis and myocardial infarction linked with causes-specific mortality                                    |
|      |                            | Chernobyl Cleanup workers (cancer, causes of death) families and children data with birth records (on irregular basis) |
|      | (In progress)              | Genomic database linked with EHRs                                                                                      |
| 9    | Finland                    | KANTA and KANSA health register linked with                                                                          |
|      |                            | Finis birth cohort 87 and 97 linked with                                                                              |
|      |                            | National HES from 1972-2017 linked with                                                                               |
|      |                            | National HIS since 1978 onwards linked with                                                                            |
| 10   | France                     | Hospital discharge linked with health insurance claims and mortality database (national health database: SNDS)       |
|      |                            | Population-based epidemiology cohorts (CONSTANCES & ELFE) linked with national health database                         |
| 11 | Germany | National health examination survey in adults linked with mortality database |
|----|---------|--------------------------------------------------------------------------------|
|    |         | National health examination survey in adults linked with health insurance claims |
|    |         | Cancer registry operated by the public health institute and included in health reporting |
|    |         | National health surveys use national and sub-national data for weighting |
|    |         | National health examination surveys use inter-metropolitan socioeconomic data for improving field work (in progress) |
|    |         | Use of socioeconomic data at the metropolitan level for small area estimation (in progress) |
|    |         | Use of real-time emergency room data for surveillance of infectious diseases (in progress in a project) |
|    |         | Linkage of data from national health surveys, health insurance data, cancer registry and other sources for national burden-of-disease calculation (in progress) |
| 12 | Greece  | No |
|    |         | Cancer registry linked with Hospital admission linked and mortality database |
|    |         | Census data linked with mortality database (one off) |
|    |         | Prescribed medication data Medical eligibility and claims data linked with income level (one off) |
| 14 | Italy   | Hospital discharge linked with mortality database and national health examination survey |
|    |         | Patient register with specific diseases linked with mortality database |
| 15 | Latvia  | Hospital discharge, primary health care, emergency care records linked with birth and mortality database |
|    | (In progress) | Patient register with specific diseases linked with mortality database |
| 16 | Lithuania | Compulsory health insurance information system (inpatient, outpatient specialized, primary emergency care) linked with causes-specific mortality database |
| 17 | Luxembourg | No |
|    |         | Health insurance claims, prescribed drugs, surgical operations, laboratory information system, radiology information system, patient administration system, outpatients attendance, pathology discharge summaries linked with birth and mortality database |
|    |         | Congenital anomalies, injuries, cancer, dementia, organ transplants registries linked with mortality database |
| 19 | The Netherlands | Health examination and interview surveys linked with mortality database |
|    |         | Health insurance claims with perinatal data |
|    |         | Cancer registry data with mortality database |
|    |         | Linkage between almost all sources by means of unique personal identification. Both within health and care services, and across other governmental areas. Big data solution in use for access and analysis. Modulation using national health registries linked with land and housing, road and transport, and other datasets. |
| 21 | Poland | Medical eligibility and claims data linked with income level (one off) |
|    | (In progress) | National health surveys linked with electronic health records |
|    |         | Hospital discharge, primary care and medical records linked with hospital registry of domestic leisure accidents, e-death certification |
|    |         | Cancer, tuberculosis, HIV and congenital anomalies registries linked with e-death certification. Hospital discharge data |
| 22 | Portugal | No |
|    |         | National registry of EHRs (Hospital discharge, general practitioner record, referrals, prescribed medications, laboratory results, diagnostic procedures medical consultations) linked with national disease-specific registries |
|    |         | National registry of EHRs linked with national registry of health care workers and health providers |
| 25 | Slovenia | National registry of EHRs linked with national registry of health care workers and health providers |
|    |         | Hospital discharge, drug prescription and perinatal health linked with mortality database |
|    |         | Hospital discharge, drug prescription and perinatal health linked with census data on educational socioeconomic variables (inequality analysis) |
|    |         | Hospital discharge, drug prescription and perinatal health linked with European Health Interview Survey |
|    |         | Mortality database linked with cancer registry |
| 26 | Serbia | National health interview survey linked with mortality database |
|    |         | Primary care data linked with drugs prescription and laboratory tests |
|    |         | National Health survey linked with cause-specific mortality data |
|    |         | All cohort studies can link their data with cause-specific mortality information through an agreement with the National Institute of statistics |
| 27 | Spain | National Patients register linked with causes-specific mortality database |
|    |         | National Patients register linked with birth database |
|    |         | National Patients register linked with dental health database |
|    |         | National Patients register linked with vaccination database |
|    |         | National Patients register linked with education, income tax, occupation, country of origin population based register |
|    |         | National health surveys (ULF/SILC), environmental health survey (MHE) and European health surveys |
interview survey (EHIS) linked national health care quality registries (each deals with a disease-specific condition)

UK-England
- UK Cancer Registry is linked with Hospital and Mortality Records
- EHRs linked with mortality database and census (demographic) database
- Cancer and diabetes registries linked with hospital and mortality records
- Scottish health interview survey linked with hospital and mortality records

UK-Scotland
- EHRs linked with each other: General and Psychiatric Hospital Stays/day cases (including intensive care/high dependency stays), Outpatient attendances, Emergency department attendances, Maternity, birth records and Neonatal Care
- EHRs linked with mortality database and census (demographic) database
- Cancer and diabetes registries linked with hospital and mortality records
- Welsh Health Survey and National Survey for Wales (interviews) linked with all of the above

UK-Wales
- EHRs linked with each other: Primary care general practice datasets linked with hospital inpatient emergency department visits, outpatient attendances, child health dataset, congenital anomaly registry, population register and laboratory results in the Secure Anonymized Information Linkage (SAIL) database www.saildatabank.com
- EHRs linked with mortality database, GIS and census (demographic) database
- Cancer, trauma and renal registries linked with all of the above
- Welsh Health Survey and National Survey for Wales (interviews) linked with all of the above
- Healthwise Wales Cohort, Millennium Cohort, Caerphilly Cohort study and UK-Biobank linked to and all of the above

EHRs linked Education Attainment records and GIS derived metrics e.g. pollution, housing quality, urban design, alcohol outlets etc.

Table 1.4: General characteristics of linked datasets in European countries in 2019 for innovative use of data sources

| S/No | General characteristics of linked datasets | Advanced N = 24 | European countries | In progress N = 2 |
|------|------------------------------------------|----------------|--------------------|------------------|
| 1    | Level of data linkage use/implementation |                |                    |                  |
|      | National level                           | 17             | AT, BE, BG, CY, CZ, EE, ES, FI, HR, LT, MT, NL, NO, SI, SK, SRB, UK (ENG, WL) | IE, LV |
|      | Sub-national level                       | 1              | IT                 |                  |
|      | Both (National and Sub-national) levels  | 3              | FR, PT, UK-SC      |                  |
|      | Metropolitan level                       | 4              | MT, PL, SI, UK-WL  |                  |
|      | All of above                             | 4              | DE, DK, NO, SW     |                  |
| 2    | Type of linkage                          |                |                    |                  |
|      | Deterministic                            | 12             | AT, CY, HR, FI, LT, MT, NL, NO, SI, SK, SRB, SW | IE, LV |
|      | Probabilistic                            | 1              | UK-SC              |                  |
|      | Combination of both (i.e., deterministic and probabilistic) | 11 | BE, CZ, DE, DK, EE, ES, FR, IT, PL, PT, UK (ENG, WL) |                  |
|      | None of the above                        | 1              | BG                 |                  |
| 3    | Current status of linked data usage      |                |                    |                  |
|      | Available and is used in routine         | 16             | AT, BE, CY, CZ, DK, EE, FI, FR, LT, MT, NL, NO, PL, SI, SW, UK (ENG, SC, WL) |                  |
|      | In progress of development               | 4              | BG, ES, HR, PT     | IE, LV |
|      | Partial in use & partial in progress of development | 2 | DE, SK |                  |
|      | Available but not in use                 | 2              | IT, SRB            |                  |
| 4    | Type of approval to access               |                |                    |                  |
|      | By law                                   | 5              | AT, C2, MT, NO, SW |                  |
|      | By ethical committee                     | 7              | BE, ES, FR, IT, NO, PT, UK (ENG, SC, WL) |                  |
|      | By register owner                        | 13             | BG, CY, ES, HR, FI, FR, IT, NO, PL, PT, SI, SK, SRB |                  |
|      | Others (i.e., depend on)                 | 7              | CZ, DE, DK, EE, LV |                  |
| S/No | Categories | Health outcomes indicators (N = 46) | European countries |
|------|------------|-------------------------------------|--------------------|
|      | Cardiovascular (N = 14) |                                |                    |
| 1    | Health characteristics | Incidence of stroke among less than and more than 65 years old | CZ, FR, LT, NL, SW, UK-WL |
|      |                        | Prevalence of stroke among less than and more than 65 years old | CZ, LT, NL, SW, UK-WL |
|      |                        | Incidence of myocardial infarction among less than and more than 65 years old | CZ, FR, LT, NL, SW, UK-WL |
|      |                        | Prevalence of myocardial infarction among less than and more than 65 years old | CZ, LT, NL, SW, UK-WL |
|      | Mortality             | Mortality due to stroke within 30-days of hospitalization | FR, IT, LV, MT, SW |
|      |                        | Mortality due to myocardial infarction within 30-days of hospitalization | CZ, FR, NO, SW |
|      |                        | Risk of mortality due to myocardial infarction at municipal level | SW |
|      |                        | Risk of mortality due to myocardial infarction at municipal level (in future) | ES |
|      | Human function and quality of life | Neurorehabilitation and functional outcomes of patients after stroke | FR, SW |
|      |                        | Neurorehabilitation and functional outcomes of patients after stroke (in progress) | NO |
|      | Life expectancy and well-being | 30-days survival following stroke | FR, NO, SW, UK-SC |
|      |                        | 90-days survival following stroke | SW, UK-WL |
|      |                        | 365-days survival following critical care for stroke | CZ, FR, LT, NL, SW, UK-WL |
|      |                        | Hospital utilization in the 365 days following intensive care discharge | CZ, FR, LT, NL, SW, UK-WL |
|      |                        | 365-days survival following stroke | FR |
|      |                        | All of above | DALYs, YLL, YLD |

Table 3.1: Description of health outcome indicators estimated using linked data across European countries in 2019
| 2 | Neurodegenerative disease (N = 6) |
|---|---|
| Health characteristics | Prevalence of Multiple Sclerosis  
| | CZ, HR, FR, LT, SW, UK-WL  
| | Prevalence of Alzheimer  
| | CZ, FR, LT, SW  
| | Incidence of Alzheimer  
| | FR, LT, SW  
| | Incidence and prevalence of Alzheimer (in progress)  
| | NO  
| Mortality | Mortality due to Parkinson  
| | FR, SW  
| | Mortality due to Dementia  
| | FR  
| 3 | Maternal and perinatal health/child health (N = 6) |
| Health characteristics | Incidence of low birth weight  
| | CY, CZ, PL, FR, NO, SW  
| | Incidence of pre-term birth rate  
| | SW  
| | Incidence of pre-term birth rate (in future)  
| | CZ, FR, NO, UK  
| | Incidence of gestational diabetes  
| | SW  
| | Incidence of gestational diabetes (in future)  
| | FR, NO  
| | Prevalence of congenital anomalies  
| | CZ, PT, LT, NO, UK-WL  
| | Emergency admissions for potentially preventable hospitalizations (PPH) between the age of 1 and 5 years  
| | UK-WL  
| Mortality | Stillbirth  
| | FR, NL, LT, NO  
| 4 | Diabetes (N = 6) |
| Health characteristics | Incidence of diabetes  
| | CZ, FR, LT, SW  
| | Prevalence of diabetes  
| | SW  
| | Incidence and prevalence of diabetes (in progress)  
| | NO  
| Mortality | Mortality due to diabetes and related risk factors  
| | BG, CZ, DE, FR  
| Human function and quality of life | Amputation rate (related complications)  
| | BE, FR, MT, LT,  
| Human function and quality of life | Number of patients with installed insulin pump during diabetes curation  
| | FR, PL, SW  
| All of above categories | DALYs, YLL, YLD  
| | SW, UK (SC, WL)  
| 5 | Suicide/Trauma/Injury (N = 7) |
| Health characteristics | Use of health care services before suicide  
| | LT, SW, UK-WL  
| | Prevalence of morbid conditions before suicide  
| | CZ, PL, FI, LT, SW  
| | Standardized prevalence of suicide  
| | FR  
| Mortality | Standardized prevalence of suicide  
| | CZ, PL, FI, LT, SW  
| | Incidence of suicide  
| | FR  
| | Incidence of injuries  
| | LT, NO, SW, UK  
| Mortality | Death rates in road accidents  
| | EE, LT, NO, SW  
| 6 | Cancer (N = 6) |
| Health characteristics | Incidence of various types of cancer  
| | CY, CZ, DE, FI, PL, NL, NO, UK-WL  
| | Prevalence of various types of cancer  
| | CZ, MT, NO, UK-WL  
| Mortality | Mortality rates due to various types of cancer  
| | CY, CZ, ES, F  
| | Life expectancy and well-being  
| | PT, LT, NO, UK-WL  
| Human function and quality of life | Scale of return to work after cancer and determining factors  
| | SW  
| Human function and quality of life | Scale of return to work after cancer and determining factors (in future)  
| | BE  
| 7 | Alcoholic liver disease and hepatic failure (N = 1) |
| Mortality | Standardized mortality ratios at 60 days and 5-years following unscheduled admissions  
| | SW, UK-WL  

Table 3.2: Description of health determinants using linked data across European countries in 2019
| S/No | Domain/Health condition | Health determinants (N = 34) | Variables can be stratified by | European countries |
|------|-------------------------|-------------------------------|---------------------------------|-------------------|
| 1    | Physical environment (N = 12) |                               |                                 |                   |
|      | Emphysema               | Air quality                   | Area of residence               | BE, UK-WL         |
|      | Injury                  | Place of injury               | Age, sex and area of residence  | CY, NO, UK-WL     |
|      |                        | Type of injury                |                                 |                   |
|      |                        | After injury hospitalized or not |                                 |                   |
|      | Parkinson               | Exposure to pesticides (i.e., agricultural activities, in vineyards, metallurgy and solvents, in textile industry) | Area of residence | FR |
| 2    | Breast cancer mortality | Industrial pollution         | Area of residence               | ES                |
|      | Adiposity               | Proximity of fast food outlets from areas of residence | Area of residence | UK (ENG, V) |
|      | Various chronic health conditions | Alcohol outlet density | Area of residence | NO, SW, (WL) |
|      | Mental health           | Presence to green-blue spaces |                                 |                   |
|      |                        | Access and visit to green-blue spaces |                                 |                   |
|      |                        | Visit to green-blue spaces   |                                 |                   |
|      |                        | Housing quality              |                                 |                   |
| 2    | Socioeconomic and environment (N = 10) |                               |                                 |                   |
|      | Multi-morbidity         | Number of single households of older people | Age and living condition | AT, NO, SV |
|      | Breast cancer mortality, injury, diabetes, cardiovascular, mental health | Sociodemographic status | Age, sex and area of residence | EE, ES, NO, PT, UK (SC, WI) |
|      |                        | Socioeconomic status         |                                 |                   |
|      |                        | Employment status            |                                 |                   |
|      |                        | Level of education achieved  |                                 |                   |
|      |                        | Deprivation index            |                                 |                   |
|      | Pre-term birth          | Maternal education to measure social disparities | | |
|      | Injury                 | Time and distance between road accident and emergency room | | |
|      |                        | Standardized absenteeism and attributable indirect costs | | |
|      | All types of cancer    | Accessibility to linear accelerators for radiotherapy | Area of residence | |
| 3    | Health behavior and life style (N = 6) | Smoking rate | Age, sex, sex, socioeconomic status and area of residence | BE, CY, MT, NL, (SC, WL) |
|      |                        | Alcohol consumption          | Age and sex                      | NO, UK-SC |
|      |                        | Physical activity            |                                 | UK-SC |
|      |                        | Dietary consumption          |                                 | UK-SC |
|      |                        | Drug use                     |                                 | NO, UK-SC |
|      | Diabetes               | Diabetes risk score          | Age, sex and area of residence   | CY, NO, SL |
| 4    | Biological/metabolic parameters (N = 3) | Self-reported BMI | Age and sex | SW, NO, UI |
|      | Obesity                |                                |                                 |                   |
|      | Diabetes               | Blindness                     | Age, sex and area of residence   | CY, FR |
|      |                        | Proteinuria                  |                                 |                   |
| 5    | Others (N=3)           | Multi-morbidity              | Age and sex                      | FR |
|      | Road accidents, neurodegenerative disease | Disability |                                | FR |
|      | Chronic health conditions | Frailty |                                | FR |

Table 3.3: Description of health intervention indicators estimated using linked data across European countries in 2019
| S/No | Categories | Domain/Health condition | Health intervention indicators (N = 23) | Member States |
|------|------------|-------------------------|----------------------------------------|---------------|
| 1    | Prevention | Maternal and perinatal health (N = 7) | Prevalence of thyroid gland examination during pregnancy | CZ            |
|      | Prevention | Low birth weight | Frequency of admission to intensive care unit | CY            |
|      | Prevention |                     | Prevalence of maternal smoking and quitting smoking during the pregnancy | FI, FR, NO, |
|      | Prevention | Pre-term birth | Percent of births in level III maternity units | FR, NO, SW   |
|      | Prevention | Perinatal mortality | Pregnant women with adequate prenatal care (number of visit/timing of initiation) | FR, NO, SW   |
|      | Prevention | Prenatal care | Screening programs as preventive check-ups during pregnancy | NO, SI, SW   |
|      | Prevention | Neural tube defect | Folic acid supplementation | PT, NO, UK   |
|      | Prevention | | | |
| 2    | Prevention | Prevention | Breast, cervical, colorectal and bowel cancer | Screening participation rates, effectiveness and evaluation | BE, EE, CZ, FI, FR, IT, SI, SW, UK-
|      | Prevention | Colorectal cancer | Frequency of surgery | CY, NO, SW   |
|      | Prevention | Colorectal cancer | Colonoscopy compliance rate | CZ, SI, NO   |
|      | Prevention | Breast cancer | Genetic screening among families (in future) | ES           |
|      | Prevention | Breast cancer | Stage distribution of detected cancer | CZ, SI, NO   |
|      | Others | All types of cancers | Re-integration in work | BE           |
| 3    | Prevention | Diabetes (N = 4) | Diabetes related complications | Foot care | BE, FR, SW |
|      | Prevention |                     | Proportion of diabetics counselled by nurse to avoid complications | SW |
|      | Prevention | | Proportion of diabetics counselled by nurse to avoid complications (in future) | CY |
|      | Prevention | Breast cancer | Amputation rate | BE, FR, SW   |
|      | Prevention | Diabetes | Percentage of diabetics with latest HbA1c above 7.0 | FR, MT, SW   |
| 4    | Prevention | Cardiovascular diseases (N = 2) | Stroke, myocardial infarction | Absolute global CVD risk assessment in primary prevention | IT |
|      | Prevention | Stroke | | Aortic aneurysm screening | SW, UK-WL    |
|      | Prevention | | | | |
| 5    | Prevention | Neurodegenerative disease (N = 2) | Multiple sclerosis | % of patients qualified for pharmacotherapy | PL |
|      | Prevention | Dementia | % of patients using neuroleptic drugs | FR |
| 6    | Prevention | Trauma/Injury (N = 1) | | Visit to primary care physicians before suicide | SI, LT |
| 7    | Prevention | COPD (Chronic Obstructive Pulmonary Disease) | % of patients with non-invasive ventilations | PL, SW |

Figures
Figure 1

Flow diagram of studies using linked data and artificial intelligence for health status monitoring to develop a questionnaire identifying various practices of data linkages across European countries in 2019
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

Additional file 3.docx
Additional file 2.docx
Additional file 5.docx
Additional file 1.docx
Additional file 4.docx