Deliverable 4.3 : Fact Sheets

July 2020

Executive summary

This report compiles 15 Fact Sheets of the Joint Action on Health Information (hereinafter referred to as InfAct) with project number 801553. InfAct is coordinated by Sciensano in Belgium and includes 40 partners in 28 countries. The project is organised through 10 work packages. The document is being delivered to the European Commission as Deliverable D4.3.

The main expected outcome of InfAct is a more sustainable research infrastructure on EU Health Information. This infrastructure will support Member States through improving the availability of comparable, robust and policy-relevant population health data and health system performance information. Through country collaboration, InfAct streamlines health information activities, reduces the data collection burden and works towards a sustainable and robust data collection in Europe that facilitates and supports country knowledge for health research and policymaking.

The Fact Sheets report on key outputs from different activities in InfAct. An overview of the Fact Sheets are provided in Table 1.

Table 1. List of Fact Sheets InfAct Outcomes

Work Package	Topic	Institution
WP5	Prioritising health information at national level	Robert Koch Institute (Germany)
WP6	Capacity building activities under European Health Examination Survey	THL (Finland)
WP6	Contributions to European Health Information Training Program (EHITP)	IHMT (Portugal)
WP6	Health Information Training Course and roadmap for sustainability	IHMT (Portugal)
WP7	Connecting health information system´s stakeholders through national nodes	Sciensano (Belgium)
WP8	Health data collection methods and procedures	ISS (Italy)
WP8	Guidance for health reports	ISS (Italy)
WP8	A sustainable ECHI shortlist	RIVM (Netherlands)
WP9	Innovative use of data sources	Santepublique France (France)
WP9	Use of artificial intelligence for public health surveillance	Santepublique France (France)
WP9	Burden of disease	Santepublique France (France)

WP9	Methodological guidelines to estimate health indicators using linked data andmachine learning techniques	Santepublique France (France)
WP9	Non health related EU databases for health surveillance. Case study industrial pollution and cancer	ISCIII (Spain)
WP9	Composite health indicators for monitoring NCD: Hospital admissions and mortality ratio	ISCIII (Spain)
WP10	Interoperability	CIPH(Croatia)/IACS (Spain)

FACT SHEET PRIORITISING HEALTH INFORMATION AT THE NATIONAL LEVEL

InfAct WP5: "Status of health information systems in European Union Member States and regions"

Key outputs

1. Duly prioritised health information is essential to ensure that health information system (HIS) indicators support public health policy action (agenda-keeping) and highlight emerging public health issues (agenda-setting).

2. Within WP5 of InfAct, Task 5.3 compiles and reviews national priority setting strategies, creating an overview of health information prioritisation across EU Member States and associated countries.

3. A two-round Delphi Survey was developed, pre-tested and conducted among representatives from EU and associated countries. Good-practice-approaches to health information development and prioritisation at national level will be identified from the experts' responses and a guidance document will be developed based on the survey findings.

4. Through uptake of this guidance in systematic establishment of national HIS, we hope to facilitate a systematic process for information prioritisation. This guidance may also inform health information prioritisation at European level for establishment of a European health information system (EU-HIS).

Background and Rationale

Health information includes data on population health, health determinants, health care systems, and health-relevant policy developments. Health information informs decision-makers, researchers and the public; it is a cornerstone of the public health action cycle[1]. Health information guides public health interventions (agenda-keeping) and points to emerging public health issues (agenda-setting). In order to fulfil these functions, health information needs to be duly prioritised, ensuring that relevant public health issues are identified and that public health interventions respond to real needs.

Interestingly, however, little can be found in the literature on health information prioritisation methods and procedures in Europe, including the selection of indicators. InfAct Task 5.3 aims to close this gap with a Delphi survey that, firstly, gathers information on national methods for health information prioritisation across the EU and associated countries. Secondly, from the information gathered in the first round, survey participants are asked to identify good-practice-approaches. The findings will be synthesised into a guidance document for prioritisation of health information for national health reporting. This guidance at the national level is a first step towards development of standards for prioritisation of health information within an EU-HIS.

Proposal

We employed an online two-round Policy Delphi survey which was distributed to EU and associated countries' representatives – mainly public health and health information experts – participating in the Joint Action InfAct.

The results of a literature review formed the basis of the Delphi survey. The 1st round contained mainly open-ended questions and was used to generate qualitative information on national health information prioritisation strategies. Full-text reponses from the 1st round were developed into closed questions for the 2nd round, focusing on prioritisation approaches, criteria and stakeholder involvement. Participants were asked to rank these questions according to the degree of "desirability", "feasibility", "importance" and "confidence", based on their expert opinion.

A total of 119 experts in 33 countries were contacted; we received 19 fully and 11 partially completed questionnaires for the 1st round of the Delphi survey. Experts from 13 countries agreed to be invited to the 2nd round; of these, six completed the 2nd survey. At the time of this writing, the analysis of both rounds is being finalised. From the results, a guidance document will be drafted, to be presented to the InfAct partners with a view to adopting a consented final version.

Recommendations for sustainability

The expected outcome of the Delphi survey is a list of good-practice-approaches to health information development and a guidance for prioritisation at national level. The document will include criteria, methods and structured prioritisation processes as well as stakeholder involvement. We also aim to draw insights into the inclusion of good-practice-approaches in the prioritisation of health information in the respective countries, as well as analyse the connection between health information and health targets, both national and international.

A guidance for prioritisation of health information for national health reporting enhances comparability of health information systems across the EU and associated countries. The guidance could be further developed into a health information prioritisation strategy at the European level for establishment of an EU-HIS.

The results will be submitted for publication in an international journal in order to affect the working practices of those developing national HIS.

References

1 Rosenbrock R (1995) Public Health als Soziale Innovation. Das Gesundheitswesen 57(3):140-144
2 World Health Organization (2014) Health in all policies: Helsinki statement. Framework for country action. WHO, Geneva

Updated Version June 2020

TECHNICAL DIALOGUE WP6

Contributions for a European Health Information Training Programme (EHITP)

Objective

The WP6.2 aimed at designing a flagship training programme to improve the member states capacities in population health and health system performance analysis and monitoring to address existing inequalities. Accordingly, the European Health Information Training Programme (EHITP) was conceptualized as an umbrella for all current and future training activities in Europe, targeting professionals working in public health and health information at national or European/international level.
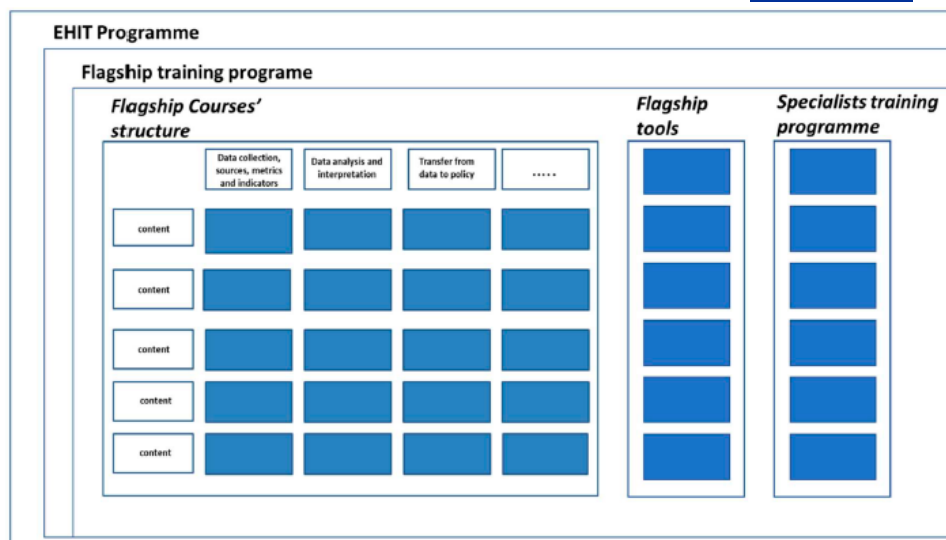
Background

Health information is a comprehensive area, in a maturing process, including data collection, data analysis and inference, indicator development, information management and translational research for developing new policies. The InfAct research outcomes, particularly on WP6, show clearly that knowledge and capacities on health information vary between European Member States (MS) and that there is a need to improve common mechanisms for strengthening the capacity to use and manage health information.

To cope with the challenges associated with strengthening Health Information capacity, health professionals require health information capabilities complying with their tasks. Nowadays, it widely recognized that most health and management functions require specific health information skills (or eSkills).

Proposal

Given that the European Health Information panorama is mainly a challenge of heterogeneous capacity rather than of lack or low capacity, the definition of a strategic plan for health information aimed to respond to the need of reducing inequities across all member states and include all relevant stakeholders and resources.

It was considered necessary to have a sustainable capacity building programme in health information that focused on the following areas: data analysis and interpretation, especially interoperability of data sources, derivation of European Core Health Indicators (ECHI) indicators and foresight/scenario analysis; transfer from data to policy, especially policy translation tools and data presentation; data collection methods, sources of data, metrics and indicators, especially issues related to health examination surveys; and data privacy and ethical issues, especially how to deal with requirements of EU General Data Protection Regulation (GDPR).

Schematic for a European Health Information Training Programme

Main results.

The EHITP proposed by WP6/InfAct aims to be an umbrella for all current and future training activities in Europe, targeting professionals working in public health and health information at national or European level. The target audience are all health related professionals in the EU MS, who can benefit from acquiring skills and competencies for addressing chronic threats to health.

Therefore, EHITP aims to meet the institutional needs of European MS to establish a competent workforce, effectively working and interacting with experts of all areas at European Level, other countries and other international organisations. It should support lifetime learning of people working in the field of public health and health information. It should be dynamic and able to respond to emerging needs in the ever-changing health information environment.

Recommendations for sustainability

1- EHITP should be a flexible structure of courses and other capacity building activities, modules and training plans, covering all the areas related to Health Information and easily tailored to tackle the different specificities.

2- Under the EHITP, MS and European Institutions should develop initiatives according to specific needs, then contributing to a European perspective of health information.

3- Modules provided by different organizations (ECDC, EMCDDA, IARC, Eurostat, OECD, WHO, etc) should be considered on the training initiatives, as well as already available academic and non-academic structures specialized training on Health Information

4- The programme must be tested through a pilot course and the evaluation of this initiative should contribute to the consolidation of a roadmap for capacity building in health information

5- More research is needed on HIS topics and their relationship with public health activities, as well as on the training of professionals for their use

FACT SHEET on
CAPACITY BUILDING ACTIVITIES UNDER EUROPEAN HEALTH EXAMINATION SURVEY (EHES)
WP6 "Strengthen EU countries' health information capacity"

Key outputs

1. Health examination surveys provide valuable information for health monitoring which cannot be obtained from other data sources such as population prevalence of hypertension, functional capacity of elderly population or vitamin D level of the population.

2. European Health Examination Survey (EHES) has prepared standardized survey protocols and guidelines and established an extensive capacity building system to support national HES organizers.

3. Information obtained through health examination surveys will support national and EU-level public health policies and development of required prevention programmes.

Background and Rationale
Health Examination Surveys (HES) are population based surveys collecting information about health and determinants of health on general population. They include questionnaires (interviewed or self-administered), physical measurements, and collection and analysis of biological samples. National HESs are an important data source for many health indicators which are not available through routine data collection (hospitalization, deaths etc.) or health interview surveys such as European Health Interview Survey (EHIS). For example, population prevalence of hypertension, functional capacity of elderly population or vitamin D level of the population is difficult to obtain from any other data source than a HES.
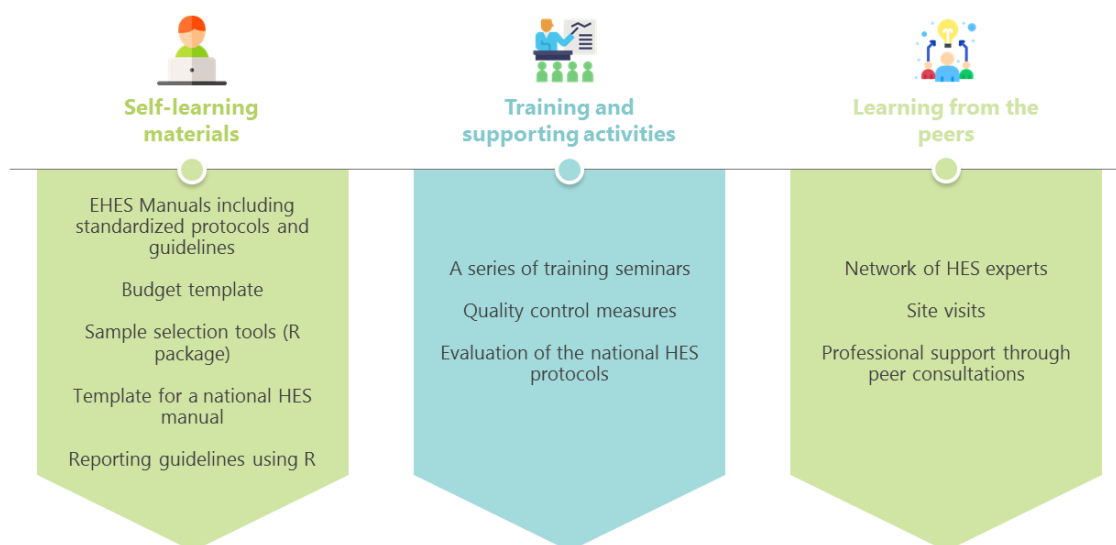
To ensure comparability of the results from these national HESs between countries and overtime within country, it is essential to use standardized protocols. To enhance the use of standardized protocols for HESs conducted in Europe, the European Health Examination Survey (EHES) initiative was established in 2009 (http://www.ehes.info/). EHES has prepared standardized survey protocols and guidelines, and set up a series of capacity building activities to support use of these standardized.

Capacity building tools in EHES
EHES capacity building activities are targeted mainly for national survey organizers following the idea 'train the trainers' and peer-support.

The EHES capacity building activities can be classified in three categories: 1) material for self-learning, 2) training and supporting activities and 3) learn from your peers.

EHES capacity building modules

Self-learning materials

EHES Manuals including standardized protocols and guidelines

Budget template

Sample selection tools (R package)

Template for a national HES manual

Reporting guidelines using R

Training and supporting activities

A series of training seminars

Quality control measures

Evaluation of the national HES protocols

Learning from the peers

Network of HES experts

Site visits

Professional support through peer consultations

Recommendations for sustainability

For cross-country comparisons, knowledge to use standardized protocols is essential. Currently, EHES network is exists but without sustainable funding. Therefore many capacity building activities have been run down or are functioning based on good will of the network members. To revive these activities, a small sustainable funding for the coordination activities would be needed.

TECHNICAL DIALOGUE

WP6

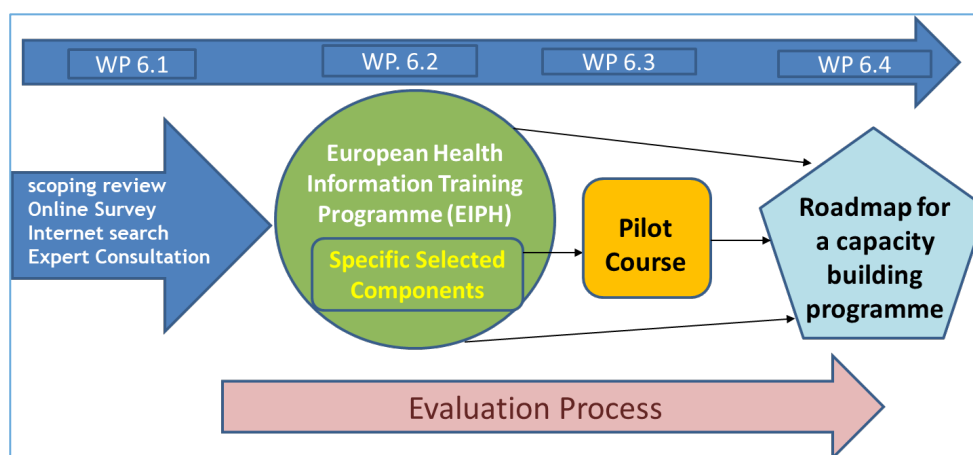**Health Information Training Course and Roadmap for Sustainability**

Objective

The objective of WP6.3 is to pilot the capacity building programme in several MSs, and provide evaluation for its implementation, contributing to the Roadmap for Sustainability,

Background

Health information is a comprehensive area, in a maturing process, including data collection, data analysis and inference, indicator development, information management and translational research for developing new policies.

As a result of research done on the InfAct Joint Action, particularly on WP6, it is clear that knowledge and capacities on health information vary among European Member States (MS) and that there is a need to improve common mechanisms for strengthening the capacity to use and manage health information.

As a result of the activities of WP6.2, a European Health Information Training Programme (EIPH) was design to be an umbrella for all current and future training activities in Europe, targeting professionals working in public health and health information at national or European/international level. This pilot course aims to test the EIPHT concept and produce recommendations for its improvement and consolidation. The combination of Tasks 6.1, 6.2 and 6.3. will enable to design a Roadmap for Health Information Sustainability in Europe.

InfAct WP 6 logical diagram

Proposal

The Course aims at addressing Fundamental Health Information tools and methods used by public health professionals and, likewise to contribute to the European Health Information Training Programme and Strategy, with a clear example of a course that could be offered by InfAct and by a Distributed Research Infrastructure on Population Health (DIPoH) in the future, contributing to improve capacity and equity in Europe.

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

The course topics will contribute for convergence in using European Methods and will be based on HI fundamentals plus innovative contributions from the InfAct work packages and experts:

The exercise will also provide additional information for further validation of the Programme. Additionally, the pilot will enable to collect important information, which will help to design the roadmap for health information equity and sustainability.

Main results.

The course consists of a week of both face-to-face and virtual sessions. This include theoretical and practical classes, group work among trainees and discussion of practical cases and projects on HIS in which trainees and trainers are involved.

Each day will be dedicated to a HIS specific topic:

- Day 1: Health information Data collection, sources, metrics and indicators:

- Day 2: Health Data analysis and interpretation:

- Day 3: Transfer from health data to policy and clinical practice:

- Day 4: Interoperability and record linkage

- Day 5: Data protection (DGPR) and ethical questions for health information

In the week before the course, the trainees will be asked to invest a few hours in the preparation of the sessions (this will be the pre-course), through readings and research and, in the following week, carry out a final work (report) for consolidation of contents and final evaluation.

Roadmap for Sustainability

The activities developed at WP6 enable to reach the following issues on sustainability of Health Information in Europe:

1- CONCEPTS: Efforts should be made to clarify concepts regarding the professions around public health activities.

2- RESEARCH: More research is needed on HIS topics and their relationship with public health activities, as well as on the training of professionals for their use. More research is needed on HIS topics and their relationship with public health activities, as well as on the training of professionals for their use

3- CAPACITY BUILDING: A sustainable capacity building programme in health information should be stablished, aiming to increase knowledge on availability and use of standardized Health Information methods, common practices within MS.

4- EUROPEAN STRATEGY: EHITP should be a flexible structure of courses and other capacity building activities, modules and training plans, covering all the areas related to Health Information easily tailored to tackle the different needs. Under the EHITP, Member States and European Institutions should develop initiatives according to specific needs and, at the same time, that contribute to a European perspective of health information.

5- EUROPEAN FLAGSHIP TRAINING: In this flagship programme, the following thematic areas should be considered as priorities: data analysis and interpretation, especially interoperability of data sources, derivation of European Core Health Indicators (ECHI) indicators and foresight/scenario analysis; transfer from data to policy, especially policy translation tools and data presentation; data collection methods, sources of data, metrics and indicators, especially issues related to health examination surveys; and data privacy and ethical issues, especially how to deal with requirements of EU General Data Protection Regulation (GDPR).

6- COLLABORATION: Collaboration among European MS and Institutions is critical for the sustainability. Training modules provided by different organizations (ECDC, EMCDDA, IARC, Eurostat, OECD, WHO, etc) should be considered on the training initiatives, as well as already available academic and non-academic structures specialized training on Health Information.

7- LEARNING: Include a cycle of learning. Guidelines and recommendations are produced and contribute to an improved version of the capacity building programme. The evaluation of this initiative will contribute to the consolidation of a roadmap for capacity building in health information

FACT SHEET
Connecting health information system's stakeholders through national nodes

---

Key outputs

1. **Background:** Health information activities are scattered in many countries. InfAct supports countries to set up national nodes that bring together key national stakeholders in the country in a systematic way and offers a structure to interact at European level.

2. **Innovative tools to improve the current EU-HIS:** InfAct identified best practices in countries with active national nodes and developed a stepwise approach to support countries to set up, define and organise a national node. A web based platform is set up to allow exchange of information on national nodes activities between countries and within the country.

3. **Most relevant results:** Countries have indicated benefits in setting up national nodes. 19 countries have provided regular updates to InfAct on their national node: 12 countries have a national node based on an existing group and 7 countries initiated first meetings in the framework of InfAct. Various stakeholders of national health information systems have shown enthusiasm in this endeavour.

4. **Feasibility of being integrated in HIS and translated into policies:** Setting up a national node brings together the key national stakeholders in a systematic way. It provides opportunities for better coordination and cooperation among stakeholders. Moreover, national nodes provide an opportunity to have a better overview of the national health data collecting agents and expertise as well as European exchange. National nodes will be an essential element in the Distributed Infrastructure on Population Health (DIPoH) and the Population Health Information Research Infrastructure (PHIRI).

Background and Rationale (What is currently known?)

In many EU Member States and associated countries, health information activities are scattered over several institutes. Regular coordination and communication among these institutes is often missing. This leads to duplication of activities, limited interoperability and linkage of data between institutes, inhibited exchange of data and lost opportunities for research or policy support.

InfAct supports European countries to overcome these barriers by setting up national nodes. The central idea is to increase communication among key players in a systematic way. The national node format is not fixed, but compatible and adaptable to the national organisation of the health information system. In some European countries, the national node function already exists in the form of a meeting organised by a coordinating institution or project. However, its function is often not clearly defined nor sustainably supported yet. InfAct provides support to countries in strengthening their function.

National nodes are foreseen to play a key role in the Distributed Infrastructure on Population Health (DIPoH) and the Population Health Information Research Infrastructure (PHIRI). The national nodes would feed DIPoH and PHIRI with relevant information, data, experts, tools and guidelines. Vice versa, they would feed relevant international information to the country.

Proposal (What tools have been used and which are the results?)

InfAct sees the national node as an opportunity to bring together regional/national stakeholders in health information in a more systematic manner. Setting up this national node brings forth discussions on core issues in health information domains that are nationally and internationally relevant for the country. Bringing together the regional/national stakeholders makes it possible to: (i) Share expertise at regional/national level; (ii) Share ongoing activities at regional/national level; (iii) Update on initiatives, meetings and expert groups at EU level and (iv) join forces for better research and policy support at national level. By enabling and strengthening national nodes, InfAct is working towards decreasing health information inequalities between countries.

InfAct reaches out to all InfAct partner countries to support them in the process of the development of the national node. To do this, InfAct initiated a national node survey to investigate the current status regarding any meetings that brought together health information stakeholders or partners at a national/regional level. More specifically, the survey collected information on how these meetings are organised, which national stakeholders are included, and what topics are discussed in these meetings in case such a meeting had already taken place. For the countries who indicate no such meetings occurred yet, the survey asked what are the potential barriers and benefits for the country to implement a national node.

The survey indicated countries saw a clear benefit of setting up a national node and countries that had already a node could witness of their added value e.g. it allows national stakeholders to work better together to respond to international request and store information centrally and accessible to all. Based on the collected experiences, InfAct developed a stepwise approach to set up a national node. The stepwise approach provides European countries with guidelines on how to set up, define, and organise a national node. Being aware that responsibilities, organisations and procedures are different in each country, the stepwise approach provided room for adjustments based on the specific situation within each country taking flexibility into account.

InfAct keeps a record on the current status of the national node in the partner countries. Countries have presented their national nodes during the General Assembly meetings and subsequent national node meetings. Opportunities for best practice exchanges and support have been organised through these regular meetings. 19 countries have provided regular updates to InfAct on their national node: 12 countries have a national node based on an existing group and 7 countries initiated first meetings in the framework of InfAct. Various stakeholders of national health information systems have shown enthusiasm in this endeavour.

Recommendations for sustainability (What should be done?)

Setting up a national node has a real benefit for countries. It brings together national data collecting agents and can optimise national data delivery. In the long term, the national node's 'spider in the web' role would also help to involve the right experts to support national capacity building, improve working towards international quality standards, and optimising the secondary use of national (and international) health data for purposes of health research, public health monitoring and health system assessment. As a result, better connected national research capacity and stronger national health information systems are facilitated by national nodes and health information inequalities are being addressed.

FACT SHEET WP8.2
A sustainable ECHI shortlist

Key outputs

1. Background: The ECHI shortlist adds value to European public health and care systems, but lacks a sustainable mechanism to maintain it.

2. Innovative tools to improve the current EU-HIS: For discussion: possible 'adoption' of the ECHI by Commission (finance and governance of the total list) and by MS (updating separate ECHI to divide the work), change the format of the list to include a more flexible and actionable part, expand the online ECHI information repository developed under Bridge Health and InfAct collectively.

3. Most relevant results: the ECHI shortlist is still considered important for EU and MS health policy by health information experts, but it needs to be updated, modernized and promoted. InfAct offers practical suggestions as to improving the lists' meta-data, content and structure, visibility and governance.

4. Feasibility of being integrated in HIS and translated into policies: A formal structure is needed to ensure the highest value to EC and MS health policies. EU entities (DG Sante, DG Estat, SGPP, WGPHS) and MS both have roles and responsibilities in this.

Background and Rationale

The European Core Health Indicators (ECHI) shortlist provides a 'snapshot' of European public health and care. It is the result of consecutive EU-wide projects starting in 1998, representing a collective MS effort and was first implemented in 2012. DG Sante maintains a webpage and an interactive tool, which is filled by Eurostat. Using ECHI to internationally compare public health aspects adds value to the national health information system. Despite the recognition of its importance by health information experts on EU and national level, there are no formal updating procedures nor is there a formal and sustainable form of governance. This puts previous efforts at risk. ECHI could be a classic example of a product for which a need was identified, that was successfully developed by MS on a project basis, but that seizes to exist because there is no infrastructure to sustain it. Therefore, InfAct aims to provide suggestions and recommendations that may benefit and improve the future of the ECHI shortlist.

Proposal

InfAct identified 4 focus areas to provide practical suggestions to:
1. **Technical updates of the metadata (documentation sheets)**
   InfAct reviewed all the ECHI documentation sheets, summarized the findings into draft recommendations and asked a group of experts to review.
2. **Modernising the content and/or structure of the list**

InfAct collected ideas for new indicators in the shortlist and developed the idea to change the structure of the shortlist and include a flexible subset to accommodate emerging information needs.

3. **Improving the visibility**

   InfAct prepared an ECHI information repository as a source of structured ECHI collective memory and input for the web portal under the RI (temporarily to be found via ECHI.eu under a website maintained by RIVM). InfAct prepared on a communication plan to increase ECHI visibility. This includes infographics, an example of which can be found on the ECHI information repository.

4. **Procedures and governance**

   InfAct drafted update procedures based on criteria that were developed by the previous ECHI projects. It also prepared a draft governance structure, with roles and responsibilities for both EU structures and MS. InfAct organised a meeting with DG Sante and ESTAT to discuss progress and possibilities to increase sustainability.

Recommendations for sustainability

In order for the ECHI to be a useful indicator set at the heart of European Health Information, it needs to be

- embedded in a sustainable infrastructure
- robust, stable and visible, and yet
- flexible to current developments

InfAct recommends that the ECHI will be embedded in formal procedures, governance and financial security, in order for it to be the useable list it was designed to be. European Commission (DG Sante and DG Estat in particular) are seen as important partners in this, with a role in securing policy relevance, technical commitment, financial sustainability and possibly legal status. The future distributed research Infrastructure on population health (DIPoH) may host the shaping and governing the ECHI, in liaison with EC and MS.

FACT SHEET GUIDANCE FOR HEALTH REPORTS

InfAct WP8: "Tools and methods for health information support"

Key outputs

1. Health reporting should provide up-to-date health-related data and information to inform policy-makers, researchers and the public. To this end, health reporting formats should be tailored to the needs and competencies of the target groups and provide comparable information of high quality.

2. The overall objective of Task 8.3 is to develop guidance for accessibility, availability and reporting of health information, including information on availability and quality of data/ indicators and the quality of reporting. Task 8.3.2 has conducted a web-based desk research to get an overview of the different formats of national health reporting and their target groups.

3. The desk research showed that health reporting practices and quality in EU Member States (MS) are heterogeneous. Currently, quality criteria and good practice examples are derived from the results of the research and translated into a guidance document on good practice for health reporting.

4. The guidance is expected to be an innovative tool for providing high-quality EU-comparable health information adequately to the targeted groups while at the same time reducing inequalities in health reporting across EU MS. The guidance will be made available to stakeholders involved in the development of national health information systems (HIS), and could be integrated into an EU HIS to enhance sustainability.

Background and Rationale

Health reporting should provide up-to-date data and information on the population's health status and its determinants, as well as on healthcare services in the countries (or regions). Establishing an information or discussion base for health policy is a key objective of health reporting ('data for action'). Policy-makers are therefore an important target group, but not the only one. Scientists and researchers, health care providers, the media and the general public are among the other addressees of health reporting.

National health reporting faces a number of important requirements. Health information should be shared in a timely manner and comparable between countries (or regions). The format and communication channel for the dissemination of health information should be tailored to the needs and competencies of the respective target groups [1].

To tackle inequalities in health reporting across EU MS and to make health information adequately accessible and available, Task 8.3.2 aims:

---

[1] Wilson PM, Petticrew M, Calnan MW et al. (2010) Disseminating research findings: what should researchers do? A systematic scoping review of conceptual frameworks. Implement Sci 5:91

(1) to prepare a comprehensive overview of the different formats of national health reporting for the dissemination of health information and their target groups.
(2) to develop a guidance document on general recommendations for good practice for health reporting in EU MS, including potential formats and target groups.
(3) to facilitate desirable and feasible criteria for creating high-quality EU-comparable public health reports.

Proposal

A web-based desk research was conducted among InfAct countries, generating a comprehensive overview of different national health reporting formats and their respective target groups. For this purpose an explorative search strategy on the status of health reporting in the EU MS has been drafted and circulated among task partners for comments, review and approval. A pre-test of the search strategy was conducted in the federal states in Germany and in partner countries and an analysis plan has been implemented for the outcome of the web-based research.
According to the findings public health reports are the most frequently used health reporting format at national level, followed by social media (e.g. Facebook, Twitter) and statistical online-databases. The general public and scientists or researchers are the most frequently stated target groups of health reporting formats. The analysis also showed that health reporting practices and quality in EU MS are heterogeneous.
The results of the web-based desk research will be used to identify good practice examples of national health reporting in different formats, based on quality criteria.
Building on the results, a guidance document with recommendations for health reporting will be developed. While the focus will be on reporting standards for public health reports, other potential formats and respective target groups will also be addressed.
The guidance document will define desirable and feasible standards for good practice while accommodating the heterogeneity of health reporting practices in the EU.

Recommendations for sustainability

There is often a gap in public health science between gaining new knowledge and its translation into practice and policy[2]. The guidance on good practice for health reporting is expected to be a sustainable tool to facilitate the generation and dissemination of health information to the targeted groups and promote access to high-quality EU-comparable information. Integrating the guidance document into health reporting training programmes could provide practical training in applying the recommendations and increase the reach. The guidance will be applicable at national as well as international level and could be integrated into an EU HIS to enhance sustainability.

---

[2] Green LW, Ottoson JM, Garcia C et al. (2009) Diffusion theory and knowledge dissemination, utilization, and integration in public health. Annu Rev Public Health 30:151-174

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

FACT SHEET Health data collection methods and procedures

InfAct WP8: "Tools and methods for health information support"

Key outputs

1. Data collection methods, metadata-reporting standards and usage of data for health monitoring (HM) and health system performance assessment (HSPA) are not uniform in Europe.

2. The objective of Task 8.1 is to identify major gaps and inequalities in health information collection methods, quality assessment, accessibility and availability procedures across EU/EEA Member States (MSs)

3. A scoping review of international organizations and selected EU research networks, and a web-based survey were carried out to identify data collection methods, availability and accessibility of HI in projects/studies performed in Europe.

4. The results of the survey highlight that evidence produced by research is not always available, comparable or usable for research purposes and policy making. Therefore, a research infrastructure providing information on standardized data collection methods and procedures and facilitating sharing and comparability of health data across EU countries is needed.

Background and Rationale

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

Inequalities and gaps in health data collection methods and in health information availability, accessibility, comparability or reusability are limiting research activities and policy making in EU countries.

To reduce gaps and inequalities of health information across MSs and to facilitate standardization of collection methods, accessibility and availability of health information, Task 8.1 aims:

• to identify European projects/studies providing HM and HSPA data;

- to summarize existing knowledge and definitions of health data, indicators, standardised data collection methods, availability and accessibility procedures covering different health data sources (e.g., population-based registries, surveys, longitudinal studies, health system performance, other administrative sources, data collected through e-health solutions, etc.) across EU/EEA MSs;

- to develop a report on health information collection methods, quality assessment, accessibility and availability procedures in and across MSs.

Proposal

The first phase of T8.1 consisted of a scoping review of international organizations (i.e., WHO-Health For All database, WHO-Health 2020 monitoring framework, WHO-Global non-communicable diseases monitoring framework, OECD, Eurostat) and selected EU research networks (i.e., European Core Health Indicators Monitoring-ECHIM, Joint Assessment Framework on Health-JAF) to identify HI data and metadata characteristics that could be used to develop a questionnaire on health data collection methods and the related harmonization and standardization procedures for health monitoring and HSPA across MSs. The identified data and metadata characteristics were then grouped into five main topics: *Source of information*, *Methodology*, *Quality*, *Data availability*, and *Data accessibility*. In the second phase of T8.1, a questionnaire based on the aforementioned topics was developed to identify data collection methods, availability and accessibility of HI in projects/studies performed in Europe. After a pilot phase, the final version of the questionnaire was administered to all representatives of the InfAct partner countries (28 MSs and 4 associated countries) through the LimeSurvey online platform.

The survey collected information on data related to 91 projects/studies from 18 EU countries. The main results of the survey show that:

- only one-third of the projects share data with EU research networks;

- less than half of the projects follow meta-data reporting standards for data description;

- less than one-third of the projects evaluate all quality criteria defined by Eurostat (i.e., relevance, accuracy, timeliness, punctuality, comparability, coherence, accessibility and clarity) and ECHO (coverage and internal reliability).

- microdata are never accessible in open access; macrodata are accessible in one-third of the projects

Basically, these results demonstrate that evidence produced by research is not always available, comparable or usable for research purposes and policy making.

Recommendations for sustainability

The survey has generated knowledge on standardized health data collection methods and procedures for health monitoring and HSPA in EU; it has also provided information on accessibility and availability of health data across EU countries. The research output will contribute to the development and the sustainability of a research infrastructure by providing information on standardized data collection methods and procedures and facilitating sharing and comparability of health data across EU countries.

FACT SHEET FOR BURDEN OF DISEASE (Ongoing activity)
WP 9 "An overview of country health profiles (i.e., GBD metrics) and a rational approach to perform a national burden of disease study in Europe"

Key outputs

1. Background: The country health profiles provide a measure of priority health conditions and risk factors, a summary breakdown of major causes, and an appreciation of health sector performance, according to the GBD methodology. There are three objectives of this study: 1. to compare the way GBD identifies top priorities in risk factors, health outcomes and health sector performance with a country's assessment based on their national health reporting, 2. to identify the potential differences in estimates due to different data sources and methods used and 3. to propose a rational approach to performing a national burden of disease study.

2. Most relevant results: There are three main results: 1. Overview of IHME produced measures of priority health conditions and risk factors across European countries. 2. Main differences in estimates due to data sources and method, and 3. A rational approach to performing a national burden of disease study (i.e., why does a country want to perform a BoD study, what are the methodologies available and what are the benefits of performing national BoD studies?)

3. Feasibility of being integrated in HIS and translated into policies: The rational approach to do a national BoD study would help to integrate the BoD approach into routine public health activities and health policies to improve the current EU-HIS (Health Information System).

Background and Rationale

The country profiles provide a measure of priority health conditions and risk factors, a summary breakdown of major causes, and an appreciation of health sector performance, according to the GBD methodology[1]. The country health profile approach highlights the usefulness and possible applications of a standardised, comprehensive methodology in Burden of Disease assessment and allows a standardised comparison with European peer countries. However, most European countries do not produce GBD metrics (YLLs, YLDs, DALYs). As a result, the national ranking of a given country based on risk factor prevalence could show that smoking is its main health problem, whereas GBD metrics (i.e., DALYs attributable to risk factors) may show that alcohol is more important.

There are three objectives of this study: 1. to compare the way GBD identifies top priorities in risk factors, health outcomes and health sector performance with a country's assessment based on their national health reporting (i.e., mortality, morbidity rates, risk factor prevalence, etc.), 2 to identify the potential differences in estimates due to different data sources and methods used by countries producing their own BoD estimates and 3. To propose a rational approach to performing a national burden of disease study.

Proposal

Using the 'standard' GBD metrics, we have produced a series of country health profiles. All charts have been produced using the same R code. All data used are publicly available, at https://vizhub.healthdata.org/gbd-compare/ and http://ghdx.healthdata.org/gbd-results-tool. Almost, 80,000 different data sources were used to produce these country health profiles. The information on data sources can be found here: http://ghdx.healthdata.org/gbd-2017/data-input-sources. After producing the charts for country health profiles of European countries, we uploaded them on an interactive website created only for 'Country Health Profiles' (https://espaces.santepubliquefrance.fr/espace_projets/Accueil/gbd). A username and password

was provided to each country, allowing to access this website of country health profiles and uploading their comments.

Expected results: We analysed and described the comments on country health profiles by taking into account the IHME estimations and the national health reporting of each country. There are three main results: 1. Overview of IHME produced measures of priority health conditions (i.e.,) and risk factors (i.e.,) across European countries. 2. Potential differences in estimates due to different data sources and methods used and 3. A rational approach to performing a national burden of disease study (i.e., why does a country want to perform a BoD study, what are methodologies available and what are the benefits for performing national BoD studies?).

The following challenges were identified by some of those countries who are performing their own BoD study when comparing GBD country health profiles with a country's own estimates: lack of comparable data, differences in data sources used, lack of information on the methods used to calculate these estimates, different methods used for age-standardized rates, differences in prevalence rates and duration parameters, different reference population[2], different life tables to estimate YLL and different methods used to redistribute garbage codes.

Recommendations for sustainability

This exercise highlighted the importance of standardised methodologies, so as to make Europe-wide BoD assessments comparable. The main intention of the BoD initiative is to integrate the BoD approach into routine public health activities and health policies to improve the current EU-HIS (Health Information System). Those countries, who are engaged in performing their national BoD studies, can provide support, guidance and recommendations to others to initiate and integrate the burden of disease approaches into their routine public health activities. At the end of three BoD workshops, we proposed the following rational approach based on a set of minimum requirements to perform a national BoD study, and its potential benefits, summarised in three main recommendations:

1. Why should a country want to perform a BoD study?
   Countries who intend to perform a national burden of disease study wish to gain a comprehensive insight into a country's health status, alongside the ability to monitor trends within a country and between countries or to produce comparable estimates at subnational and national levels.

2. What methodologies are available?
   IHME GBD manual 2020, WHO 2001 a practical guide on national BoD studies[3], some examples of national BoD studies and resource links to BoD related published studies, reports, websites, etc.

3. What are the benefits of performing national BoD studies?
   - Direct benefits: Health improvement - to inform health policy with a global and integrated approach for population health, rational allocation of resources, etc.
   - Indirect benefits: Capacity building - to build up local capacities and expertise, to use the best available data for your country.
   - Indirect benefits: Quality of HIS - to appraise and improve completeness and quality of the country's health information system.

References

1. IHME. Country Health Profiles: http://www.healthdata.org/results/country-profiles.
2. Wyper GMA, Grant I, Fletcher E, McCartney G, Fischbacher C, Stockton DL. How do world and European standard populations impact burden of disease studies? A case study of disability-adjusted life years (DALYs) in Scotland. *Archives of Public Health.* 2020;78(1):1.
3. WHO. National Burden of Disease studies: A Practical Guide: https://www.who.int/healthinfo/nationalburdenofdiseasemanual.pdf. 2001.

FACT SHEET INDUSTRIAL POLLUTION AND CANCER
WP9 "Innovation in health information for public health policy development"

Case study: Use of information from non-health related UE databases for health surveillance: the inclusion of E-PRTR data into spatial mortality and morbidity analyses

Key outputs

1. The integration of non-health related UE databases is a feasible strategy to enrich and develop health surveillance. The design of ad-hoc instruments for this purpose can facilitate the use of these data as well as the comparability of the results among EU countries.

2. This case-study presents a practical example, focused on the use of the data included in the European Pollutant Release and Transfer Register (E-PRTR), which allows estimating industrial pollution exposure, for cancer surveillance.

3. "*En-risk*", an easy-to-use java/web application tool, allows merging, at country level, a) the information of the location of industrial facilities included in the European Pollutant Release and Transfer Register (E-PRTR); and b) the municipal mortality or morbidity data. Afterwards, it performs an exploratory spatial analysis of association between them by type of exposure

Background and Rationale

Being able to combine health information with environmental health determinants is very important, both for surveillance or epidemiological monitoring and for risk studies in health. Within the European Union, there are many non-health data that can be used in this context. However, due to the heterogeneity in the availability and in the formats of this data, its integration with health data is difficult and represents an important challenge, which, in many cases, is complex and require specific expertise.

A good example of a potentially useful source of significant environmental data that might be relevant for health is the European Pollutant Release and Transfer Register (E-PRTR), which allows estimating exposure to industrial pollution, a very relevant environmental risk factor from a public health point of view (1). The register, maintained by the European Environmental Agency, contains annual data on more than 30,000 industrial facilities that reported emissions over a determined threshold of any of the selected 91 pollutants. The list of industries comprises 65 economic activities within 9 industrial sectors. For each facility, E-PRTR provides information on type of activity, geographical location and emissions of polluting substances. These data can be freely downloaded from the website of the registry. E-PRTR was established through Regulation (EC) No 166/2006, and covers 28 EU Member States as well as Iceland, Liechtenstein, Norway, Serbia and Switzerland

Proposal
In this case study, we are piloting "*En-risk*", an easy-to-use java/web interactive application tool that merges, at country level, the information of The European Pollutant Release and Transfer

Register (E-PRTR) and the municipal mortality or morbidity data to perform an exploratory spatial analysis of association between them by type of industrial facility. The application works in the user computer. It downloads the geographic coordinates for each facility from the official web of the E-PRTR, while health data can be directly loaded into "*En-risk"* by the user. This way, health information is always stored and managed in the computer of the user in order to guarantee data protection.

The application needs web connection (but could be optional) as well as the following minimum data:

- **Shapefile** (cartography) of the country (spatial unit = municipality)

- Annual **observed deaths** (for mortality) or **cases** (for morbidity) and **population figures broken down by age groups (18) and by sex** per municipality

- Optional information that could also be loaded by the users: social and economic environment information at municipal level.

With this information the application directly calculates

a) The expected number of deaths or of cases of the selected disease, using as reference the rates by age group and sex for the whole country.

b) The distance from the municipal centroids (information obtained from the shapefile) to the location of all the industrial facilities included in the E-PRTR. These distances allow classifying municipalities as exposed or not exposed to industrial pollution, according with the definitions included in the methodological annex.

With these elements, and thanks to the extended expertise of the research team in this field, *En-risk* performs a complex spatial association screening analysis that allows to evaluate whether there is any excess mortality/morbidity in those municipalities exposed to industrial pollution compared to those not exposed, globally and by industrial sectors. If the user has loaded additional information (social and economic environment information at municipal level), the analysis could be also performed considering them as possible confounding factors.

The final output of the application will include: 1) a table or forest plot of Relative Risk of mortality due to exposure to industrial pollution by industrial sector and disease analyzed; 2) a standard database adding the environmental exposure to the health data provided, that might be used for further analyses or allow looking to the spatial distribution of the exposure. However, other types of outputs can be developed. For example, ranking of municipalities with excess of risk, spatial cluster analysis, etc.

The use of the application does not require statistical knowledge, although the interpretation of the results clearly needs public health expertise. In addition, if this initial screening indicates the presence of any health problem linked to residential proximity to industrial pollution in any country, it should be followed by ad-hoc studies to deepen into it.

Recommendations for sustainability

The formulation of the European Directive on Integrated Pollution Prevention and Control (IPPC) and the creation of the EPRTR enable Member States to incorporate information of industrial pollution sources from E-PRTR into health information system, which is homogeneous and comparable among European countries. *En-risk* facilitates the study of the relationship between pollutant groups, type of industrial sector and health effects such as cancer all around all Europe. It can be used by public health services to identify health problems and to point to key policy interventions to reduce the impact of industrial pollution on health. In addition, the same approach, handy and cheap, can be applied to other geographically-based European environmental databases. Finally, its sustainability is clear because is a normative tool that might improve interoperability of health information systems with non-health data, which would be included in machine learning algorithms in the future.

References

(1) Fernández-Navarro, P.; García-Pérez, J.; Ramis, R.; Boldo, E.; López-Abente, G. Industrial pollution and cancer in Spain: An important public health issue. Environ. Res. 2017, 159, 555–563.

FACT SHEET: HOSPITAL ADMISSIONS AND MORTALITY RATIO: A COMPOSITE HEALTH INDICATOR FOR MONITORING NON-COMMUNICABLE DISEASES

WP9 "Innovation in health information for public health policy development"

> **Case study:** Development of composite indicators to monitor the burden of disease and health system performance at population level

Key outputs

1. Morbidity and mortality rates are the most used epidemiological indicators to describe the health status of a population. Although these indicators are often correlated, there are also some differences between them that should be considered in the epidemiological analysis.
2. Combined morbidity and mortality analyses could provide complementary information on disease patterns that might not arise from the independent analysis of each indicator.
3. The usefulness of a composite indicator (defined as the ratio of hospital morbidity to mortality rates in monitoring non-communicable diseases) is explored.

Background and Rationale

Worldwide, Non Communicable Diseases (NCD) are one of the major health challenges of the 21st century. Indeed, the World Health Organization (WHO) has set as a goal a reduction of 25% in global mortality associated with cardiovascular diseases, cancer, diabetes and chronic obstructive respiratory diseases by the year 2030.

In the assessment of the health status of any population, the burden of disease is usually estimated from the analysis of morbidity and mortality rates, which are the basic epidemiological indicators in public health surveillance. Although morbidity is usually correlated with mortality, there are also some discrepancies between these indicators that have been described in cardiovascular diseases and in some types of cancer. This discordance could be related to local and regional differences in population´s wealth, the degree of prevention and control of risk factors, or the availability of health resources.

Therefore, the analysis of the epidemiological patterns of NCD should include an integrated study of morbidity and mortality, describing their geographic variability and, if detected, examining their causes. This indicator would provide complementary information in the analysis of the population´s health status, using sources of information and procedures already implemented in public health surveillance systems.

Proposal

This study analyses the ratio of age-adjusted hospital morbidity and mortality rates for the following NCDs: ischemic heart disease (IHD), cerebrovascular disease (CVD), chronic obstructive pulmonary disease (COPD), and prostate, breast and lung cancer. This composite indicator has been developed from the Hospital Morbidity Survey and the Death Statistics in all

of Spain and its provinces, information which is provided by the National Statistics Institute (INE, Spanish acronym). The minimum data necessary for its construction were:

a. Hospital morbidity occurred in 2016: number of hospital admissions according to sex, age, main diagnosis, type of hospital admission, and province of residence.
b. Deaths occurred in 2016: number of deaths according to sex, age, main cause of death and province of residence.

With such information for each disease, the morbidity and mortality rates were calculated in men and women for each province per 100,000 inhabitants, adjusted by age (European standard population of 2013). Subsequently, the ratio of the morbidity and mortality rate (hereinafter HMR), and the HMR male/female ratio were calculated. In the case of IHD, CVD and COPD, these analyses were repeated by selecting only emergency admissions. For the geographical representation of the 50 provinces and 2 autonomous cities, quintiles of the three indicators (morbidity, mortality and HMR) were estimated. All analyses were performed with Stata® v.15, using the spmap module for spatial data representation.

Summary of results

Figure 1 shows the variability of HMR in men and women for the health problems studied. In cardiorespiratory diseases, COPD stood out, with an average of 7 admissions per year in men and 9 to 10 in women for each deceased, with great variability between provinces. IHD and CVD showed more homogeneous HMR, but unlike COPD men had higher values than women. Regarding to cancer, lung cancer was the most homogeneous in both sexes, while breast cancer showed the highest values and the greatest variability.

Geographical patterns are observed, depending on the diseases involved. Using CVDs as a case study, the distribution of HMR by province and sex is shown in figures 2 and 3. The southern provinces of the peninsula (the region of Andalusia) concentrated the lowest values so that reported higher proportional mortality (figure 2). In addition, large Pearson's correlation coefficients, r= -0.69 (p<0.001) and -0.71 (p<0.001) were observed in both men and women (Figure 3). This association was also observed for IHD in both sexes, COPD in men and breast and lung cancer in women, albeit their slopes were less steep than CVD.

Recommendations for sustainability

The analysis of this indicator showed important geographic variability that should be further explored to identify potential associated factors. There is also a need of examining specific case-management approach at hospital level that could explain the trends observed in the HMR and mortality in Spain. This information might increase the knowledge of the epidemiological distribution of these diseases, providing additional information to the separated analysis of morbidity and mortality.

The HMR is a tool that uses standardized methods and is based on routine data sources and traditional analytical procedures in public health surveillance systems. This indicator would allow a better understanding of regional variability between and within countries, and can also be useful for health planning and prevention.

Annex

Figure 1. Distribution and variability of hospital morbidity and mortality ratio (HMR) values according to selected diseases.
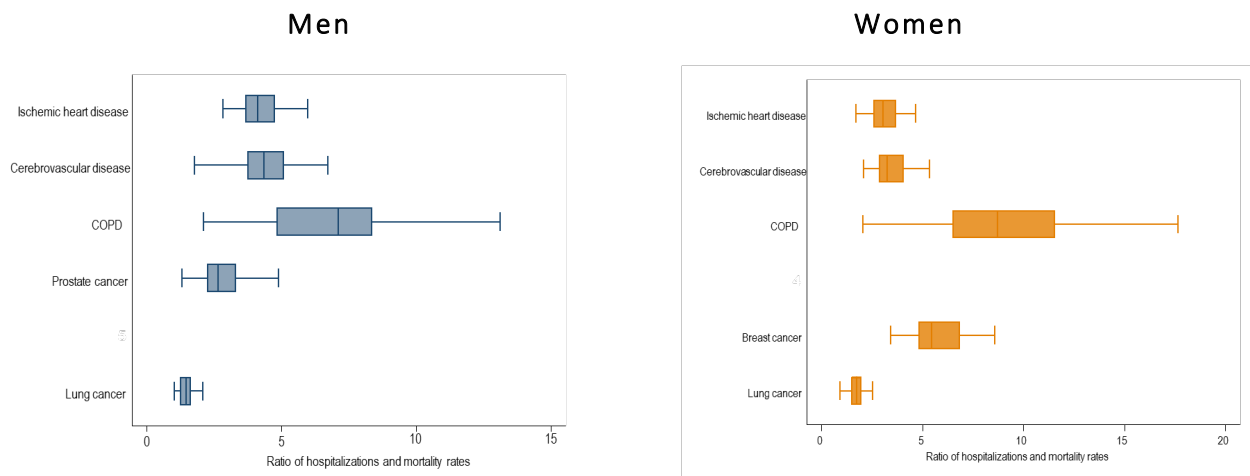

Figure 2. Geographical distribution of hospital morbidity and mortality ratio (HMR) for cerebrovascular disease by sex, 2016.
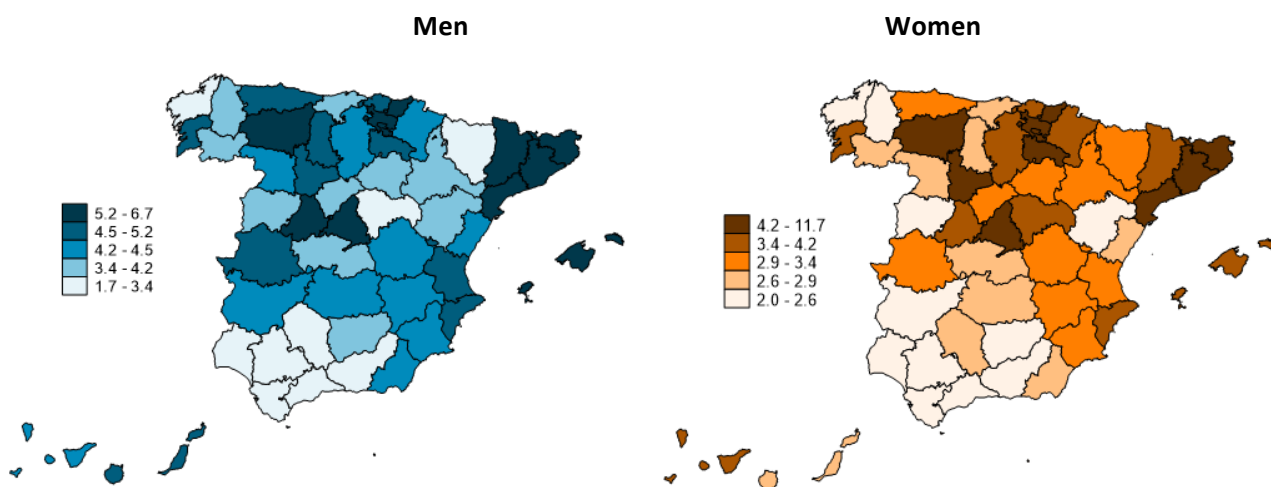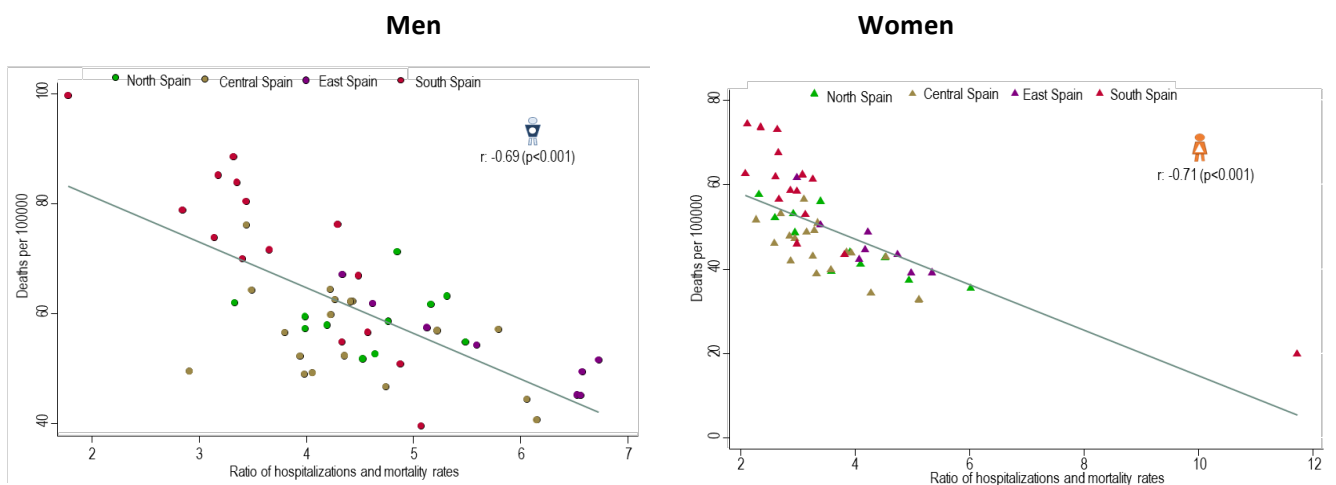

Figure 3. Correlation and linear fit between HMR and age-adjusted mortality rates for cerebrovascular diseases by sex, 2016.

FACT SHEET FOR INNOVATIVE USE OF DATA SOURCES
WP9 "Innovative use of data sources: A cross-sectional study of data linkage and artificial intelligence practices across European countries"

Key outputs

1. Background: The availability of data generated from different sources is increasing with the possibility to link these data sources with each other. However, linked administrative data can be complex to use and may require advanced expertise and skills in statistical analysis.

2. Innovative tools to improve the current EU-HIS: Use of data linkage of different administrative sources and AI to analyze large datasets are innovative tools, which are essential to improve the current EU-HIS.

3. Most relevant results: The use of data linkage has been integrated in routine public health activities among majority of European countries but only a few use AI. Using linked data, 46 health outcome indicators, 34 health determinants and 23 health intervention indicators were estimated in routine. The complex data regulation laws, lack of human resources, skills and problems with data governance, were reported by European countries as obstacles to routine data linkage for public health surveillance and research purposes.

4. Feasibility of being integrated in HIS and translated into policies: A sustainable national HIS and a robust data governance framework allowing to link different data sources are essential to support evidence-informed health policy development.

Background and Rationale

The availability of data generated from different sources is increasing with the possibility to link these data sources with each other. 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 artificial intelligence (AI) in routine public health activities, to identify the related estimated health indicators (i.e., outcome and intervention indicators) and health determinants of non-communicable diseases and the obstacles to linking different data sources.

Proposal

We performed a survey across European countries to explore the current practices applied by national institutes of public health, health information and statistics for innovative use of data sources (i.e., the use of data linkage and/or AI)[1].
The use of data linkage and AI at national institutes of public health, health information and statistics in Europe varies[1]. 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, health information and statistics. Using linked data, 46 health outcome indicators, 34

health determinants and 23 health intervention indicators were estimated in routine. The complex data regulation laws, lack of human resources, skills and problems with data governance, were reported by European countries as obstacles to routine data linkage for public health surveillance and research.

Recommendations for sustainability

To address the above mentioned obstacles and to increase the uptake of innovative and high-performance technologies in public health activities, 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 [2], 2. Specific mandates to ensure data availability/access/capture and safe storage should be an integral part of a national/regional health information system, and 3. Differences in the implementation and interpretation of the EU-GDPR (General Data Protection Regulations) and additional national regulations should be mapped and if possible harmonized across EU-MSs [3].

B. Technical aspects: 4. More collaborations and partnerships should be encouraged to build up capacities for using new health information related technologies, to share new methods, skills, experiences and data for comparative research studies among EU national institutes of public health, health information and statistics;

C. Data Governance, 5. Initiatives to strengthen national health information infrastructure should be encouraged.

D. Organizational and structural aspects, 6. Ministries of health and research from European countries should provide their support (i.e., financial and political) for the development of integrated national health data hubs/data platforms to strengthen the national health information infrastructure.

References

1. Haneef R, Delnord M, Vernay M, et al. Innovative use of data sources: a cross-sectional study of data linkage and artificial intelligence practices across European countries. *Archives of Public Health.* 2020;78(1):55.
2. 2020 C. Data Protection and Artificial Intelligence: https://edps.europa.eu/data-protection/our-work/subjects/artificial-intelligence_en.
3. 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.

FACT SHEET FOR METHODOLOGICAL GUIDELINES (Ongoing activity)
WP9 "Methodological guidelines to estimate health indicators using linked data and machine learning techniques"

Key outputs

1. Background: The estimation of health indicators from linked data and the application of machine learning techniques is challenging and may require advanced expertise and skills in statistical analysis. There is a need for methodological guidelines, which could systematically guide MSs for using linked data and machine learning techniques to estimate health indicators for public health research.

2. Innovative tools to improve the current EU-HIS: These guidelines were developed for the first time to adopt, develop and compare new methods/techniques using linked data and machine learning techniques. In turn, this would help to improve the methodological approaches for public health research.

3. Most relevant results: These guidelines contain the following seven important contents: 1. Rational and objective of the study (i.e., research question), 2. Study design, 3. Study population/sample, 4. Linked data sources, 5. Study outcomes, 6. Data preparation and 7. Data analysis. These aspects would be described by using the examples of methodological studies.

4. Feasibility of being integrated in HIS and translated into policies: These guidelines would support to improve the quality and comparability of health information. The estimated health indicators would guide policy process.

Background and Rationale

The capacity to use data linkage and/or the use of artificial intelligence to estimate and predict health indicators varies across EU-MSs (European Member States)[1]. However, the estimation of health indicators from linked administrative data is challenging due to several reasons such as variability in data sources and data collection methods, availability of a large number of variables, lack of skills and capacity to link and analyze big data[2]. To our knowledge, there are no methodological guidelines available, which could systematically guide MSs for using linked data and machine learning techniques to estimate health indicators. Therefore, the InfAct project has proposed to develop these guidelines, which could guide those MSs who are planning to estimate health indicators using linked data and artificial intelligence with new methods/techniques. The main objective of this study is to develop the methodological guidelines for studies to guide European countries using linked data and artificial intelligence with new methods/techniques.

Proposal

We have performed four following steps systematically to develop the methodological guidelines: 1. scientific literature review; 2. development of generic method; 3.

identification of inspiring examples or best practices from European countries; and 4. validation by a panel of experts.

Expected Results: These guidelines contain the following seven important contents: 1. Rational and objective of the study (i.e., research question), 2. Rational for the selection of a study design, 3. Selection of study population/sample, 4. Linked data sources available, 5. Defining the study outcomes, 6. Data preparation and 7. Data analysis. We have described these aspects with examples of different methodological studies.

Recommendations for sustainability

These guidelines aim to adopt, develop and compare new methods/techniques using linked data and machine learning techniques for public health research studies.

We proposed the following recommendations based on these guidelines:
- Rational selection of the study design using linked data is important and may avoid certain methodological limitations.
- Standard methods for data collection should be implemented in a HIS.
- Routinely data collected from various administrative sources should improve their quality concerning to the completeness of information.
- Data related to employment, education, occupation and socioeconomic status should be readily available/accessible for analysis related to the health status.
- Specific mandates to ensure data availability/access/capture and safe storage should be an integral part of a national/regional health information system[1]
- More collaborations among Member States for an exchange of inspiring examples/best practices in using linked data and machine-learning approaches are needed in the future among European countries.
- Joint country studies on using machine-learning techniques for public health research are needed.
- Better approaches to translate estimated health indicators for health policy are required.

Such guidelines need to be revised after inputs from experts.

References

1. Haneef R, Delnord M, Vernay M, et al. Innovative use of data sources: a cross-sectional study of data linkage and artificial intelligence practices across European countries. *Archives of Public Health.* 2020;78(1):55.
2. 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.

FACT SHEET FOR USE OF ARTIFICIAL INTELLIGENCE FOR PUBLIC HEALTH SURVEILLANCE
WP9 "A case study to develop a Machine Learning (ML)-algorithm to predict the incidence of Diabetes Mellitus and a summary of inspiring examples using linked data and ML techniques"

Key outputs

1. Background: The possibility to link different data sources with each other and the use of artificial intelligence to analyze large datasets are increasing in healthcare. 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 develop a generic approach to predict a health outcome from linked data set using machine-learning technique and to identify inspiring examples applying these innovative techniques in public health across European countries.

2. Innovative tools to improve the current EU-HIS: The innovative techniques (i.e., data linkage and/or artificial intelligence) have several advantages, which can improve the current EU-HIS by enhancing completeness and comprehensiveness of information to guide health policy process, by reducing the dimensionality of large datasets and more efficient analysis of large datasets with high precision.

3. Most relevant results: The generic ML-algorithm was a linear discriminant model based on 23 variables related to the biological tests, drugs, medical acts and hospitalization without a procedure over last two years to predict the incidence of diabetes. This algorithm has a sensitivity of 62%, a specificity of 67% and an accuracy of 67% [95%CI: 0.66 – 0.68]. We have identified 16 studies (12 studies related to data linkage, 2 studies applied machine learning and 2 studies used both data linkage and machine learning approaches) as inspiring examples from ten European countries.

4. Feasibility of being integrated in HIS and translated into policies: These results would support countries to learn from each other, to develop, adopt and integrate these innovative approaches to estimating health indicators and to translate those evidence into policy.

Background and Rationale

The possibility to link different data sources with each other and the use of artificial intelligence to analyze large datasets are increasing in healthcare. These innovative techniques (i.e., data linkage and/or artificial intelligence) have several advantages such as data linkage improves completeness and comprehensiveness of information to guide health policy process, whereas the artificial intelligence allows handling data with a large number of dimensions (features) and units (feature vectors) more efficiently with high precision. However, linked administrative data can be complex to use and may require advanced expertise and skills in statistical analysis. The capacity to use data linkage and/or the use of artificial intelligence to estimate and predict health indicators varies across EU-MSs (European Member States). The main objectives of this study were to develop a generic approach to predict a health outcome from linked data set using machine-learning technique and to identify inspiring examples applying these innovative techniques in public health across European countries.

Proposal

To develop the generic approach, we adopted a supervised machine learning approach[1]. Following steps were performed: i. selection of final data set, ii. case/target definition, iii. coding features/variables for a given window of time, iv. split final data into training and test data sets, v. features/variables selection, vi. training model/algorithm, vii. validation of model/algorithm with test data set and viii. selection of the model/algorithm.

The final data set used to develop the ML-algorithm included 44,659 participants and 3468 SNDS variables were coded similarly. Only 23 were selected to train different algorithms. The final algorithm was Linear Discriminant Analysis (LDA) model based on the number of reimbursements of 23 variables related to biological tests, drugs, medical acts and hospitalization without a procedure over last two years to predict the incidence of diabetes. This algorithm has a sensitivity of 62%, a specificity of 67% and an accuracy of 67% [95% CI: 0.66 – 0.68].

We have identified 16 studies (12 studies related to data linkage, 2 studies applied machine learning and 2 studies used both data linkage and machine learning approaches) as inspiring examples from ten European countries[2]. These studies covered 14 different domains of public health. Some of these studies applied classical statistical methods such as multilevel linear regression and some of these studies used artificial intelligence such as machine learning techniques. These studies highlighted that different data collection method, lacking completeness of information or inaccessibility to certain information makes challenging to analysing large linked datasets. Using linked data and AI, the methodological and data analysis aspects can be improved. The results of these studies are used to improve public health surveillance, developing prevention strategies, evaluating health care services and guiding health policy process.

Recommendations for sustainability

We recommend further research to improve the performance of this algorithm to applying on SNDS and to predict the type II diabetes cases in real-time data. More research is needed using various MLTs to predict the incidence of various health conditions by taking into account various determinants (i.e., health and non-health) for improved public health surveillance.

These inspiring examples would support countries to share different experiences and to learn from each other. Furthermore, these examples would help countries to develop, adopt and integrate innovative approaches using data linkage and artificial intelligence to estimating health indicators. These examples also allow comparing various approaches used for innovative use of health information across MSs. These inspiring examples would support to develop the methodological guidelines, which would allow estimating health indicators using linked data and artificial intelligence. Eventually, the evidence produced by using innovative techniques would guide policymaker to make better decisions.

References:

1. Haneef R., et al. "Use of artificial intelligence for public health surveillance: a case study to develop a machine-learning algorithm to predict the incidence of Diabetes Mellitus", (manuscript in preparation).
2. "Inspiring examples from European countries: Implications for innovative use of data sources", InfAct Deliverable 9.2.2, 2020.

FACT SHEET INTEROPERABILITY
WP10

Key outputs

1. Interoperability is a key factor for establishing a holistic European health data infrastructure able to translate data, information and knowledge.

2. Semi-structured in-depth interviews were conducted with key opinion leaders from different European cross-border projects that dealt with sharing, linking and managing health data. The development of a distributed infrastructure was the pillar of the European Interoperability Framework (EIF) and the FAIR principles was piloted.

3. Qualitative analysis results of our semi-structured in-depth interviews enable us to better understand the enablers and the barriers to the cross-border linkage and sharing of health data through four interoperability layers (legal, organisational, semantic and technical). Feasibility of complying with GDPR and Ethical principles was assessed, adapting to the organizational specificities of each data hub, assuring semantic interoperability across hubs and developing technological interoperability and feasibility of the development of the FAIR principles has been also tested.

4. Recommendations and publications of our results which are derived from key opinion leaders from different European cross-border projects that dealt with sharing, linking and managing health data will enable better optimization and utilization of health information systems across Europe and will facilitate development of health information and research infrastructure based on cumulative experiences and know-hows from key opinion leaders. Pilot cases on federated infrastructure and recommendations that derive from them will facilitate the implementation and development of federated infrastructure and will facilitate the deployment of the distributed solutions for data linkage, data extraction, data analysis and data reporting.

Background and Rationale

Health data allows us to create a holistic view of the overall healthcare system, enables us to conduct research and provides the basics for creating health policies. Every year there is more and more health data being generated on the European level, yet not all of that data is being utilized to its fullest potential because of the lack of interoperability. Interoperability is a key factor for establishing a holistic European health data infrastructure able to translate data, information and knowledge. The objective of our package was to map out and assess cross-national inspirational experiences on data reuse for both public health research and monitoring initiatives as well as to pilot interoperability in a number of topics relevant to public health research, using a variety of data sources from a number of locations.

Proposal

Semi-structured in-depth interviews were conducted with key opinion leaders from different European cross-border projects that dealt with sharing, linking and managing health data with a goal to better understand the enablers and the barriers to the cross-border linkage and sharing of health data through four interoperability layers (legal,

organisational, semantic and technical). Semi-structured interviews were conducted via webcam using GoToMeeting software, recorded and transcribed. Transcripts of the semi-structured interviews were analyzed qualitatively by framework analysis. The development of a distributed infrastructure was the pillar of the European Interoperability Framework (EIF) and the FAIR principles was piloted.

Preliminary results

Achieving interoperability with health data is a long process with many obstacles. Most key opinion leaders emphasize legal and semantic interoperability layer as a main barrier, while key opinion leaders no longer see technical interoperability as a barrier unless practicing physicians and patients are involved. Organisational interoperability was less understood by key opinion leaders, we are yet to analyze if it was due to our interview structure or duo to lack of understanding of organisational interoperability.

Other barriers, which were emphasized by key opinion leaders, were lack of funding, differences in health data in countries with decentralized governments and different interpretations of the GDPR, which varied between countries, between different regions of a country and between different institutions. Other enablers, which were emphasized by key opinion leaders, were univocal health data in countries with centralized governments, pre-existing legislation for a specific topic in certain countries and continuation to a work done by pre-existing project.

Feasibility of complying with GDPR and Ethical principles was assessed, adapting to the organizational specificities of each data hub, assuring semantic interoperability across hubs and developing technological interoperability and feasibility of the development of the FAIR principles has been also tested.

Recommendations for sustainability

The results of qualitative analysis of our semi-structured in-depth interviews will serve as basis for publishing recommendations and publications which are derived from key opinion leaders from different European cross-border projects that dealt with sharing, linking and managing health data which will enable better optimization and utilization of health information systems across Europe and will facilitate development of health information and research infrastructure. The recommendations and publications have the advantage of sustainability because they will serve as basis for health policies and for improving interoperability of health information systems. Pilot cases on federated infrastructure and recommendations that derive from them will facilitate the implementation and development of federated infrastructure and will facilitate the deployment of the distributed solutions for data linkage, data extraction, data analysis and data reporting.