

MORBIDITY STATISTICS

2019 Pilot Data Collection Belgium Final Report

NIENKE SCHUTTE • LAURENS RAES • BRECHT DEVLEESSCHAUWER

WHO WE ARE

SCIENSANO can count on more than 700 staff members who commit themselves, day after day, to achieving our motto: Healthy all life long. As our name suggests, science and health are central to our mission. Sciensano's strength and uniqueness lie within the holistic and multidisciplinary approach to health. More particularly we focus on the close and indissoluble interconnection between human and animal health and their environment (the "One health" concept). By combining different research perspectives within this framework, Sciensano contributes in a unique way to everybody's health.

For this, Sciensano builds on the more than 100 years of scientific expertise of the former Veterinary and Agrochemical Research Centre (CODA-CERVA) and the ex-Scientific Institute of Public Health (WIV-ISP).

Sciensano

Epidemiology and public health • Lifestyle and chronic diseases

December 2020 • Brussels • Belgium
Internal reference number: D/2020/14.440/86



Nienke Schutte



Laurens Raes



Brecht Devleesschauwer

Contact person: Brecht Devleesschauwer • T+32 2 642 50 35 •
brecht.devleesschauwer@sciensano.be

Funding: This project was funded by Eurostat, the statistical office of the European Union (call ESTAT-PAS-F-MORBIDITY-2018-1, grant agreement number: 847046 – 2018-BE-MORBIDITY).

Please cite as: Schutte N, Raes L, Devleesschauwer B. Morbidity Statistics. 2019 Pilot Data Collection Belgium, Final Report. Brussels, Belgium: Sciensano; 2020. 52pp. Report number: D/2020/14.440/86.

LIST OF ABBREVIATIONS

BE	Belgium
BHIS	Belgian Health Interview Survey
EPS	Echantillon Permanent - Permanente Steekproef
EMR	Electronic Medical Record
EU	European Union
FPS	Federal Public Service
GAM	Generalised Additive Model
GP	General Practitioner
HDD	Hospital Discharge Data
ICPC	International Classification of Primary Care
IMA-AIM	Intermutualistisch Agentschap / Agence Intermutualiste
IVC-CSI	Informatieveiligheidscomité / Comité de Sécurité de l'Information
MoH	Ministry of Health
MZG-RHM-MHD	Minimale Ziekenhuisgegevens / Résumé Hospitalier Minimum / Minimal Hospital Data
NHSD	National Hospital Stay Database
PH	Public Health
RIZIV-INAMI	Rijksinstituut voor Ziekte- en Invaliditeitsverzekering / Institut National d'Assurance Maladie-Invalidité
SGP	Sentinel Network of General Practitioners
TCT	Technische Cel / Cellule Technique

CONTENTS

List of abbreviations	5
1. Key Points	7
2. Executive Summary	8
3. Introduction	9
4. Procedures and Methods	11
4.1. Stakeholders Involved	11
4.2. Collaborations	13
4.3. Data Source Selection.....	16
5. Development and Successes	18
5.1. Acquired Data.....	18
5.2. Analysis Infrastructure	21
5.3. Corrections	24
6. Encountered Difficulties	25
6.1. Acquiring Data Access	25
6.2. Definition and Data Compliance	25
6.3. Data Quality.....	27
7. Recommendations	32
7.1. Indicator Shortlist.....	32
7.2. Data Quality Evaluation System	33
7.3. The Need for a National Health Information System	34
7.4. The Need for Nationally Representative Primary Care Data	34
Acknowledgements	36
References	37
Annexes	38
A. Data Source Selection	38
B. Meetings with External Stakeholders	51

1. Key Points

Key points:

- The project required the involvement of multiple stakeholders in Belgium. Some of these collaborations were initiated while paving the way for enhanced future data exchanges and collaborations within other projects.
- Methodologies and tools were developed to increase the reproducibility of the analyses and to decrease the risk of errors due to miscoding or the collection of variables and facilitating future work.
- We propose a reduced set of indicators for inclusion, as the validity of some indicators is questioned and there are some concerns regarding the compatibility and description of some of the proposed definitions.
- We propose an easy to read scoring system to evaluate the quality and comparability of the different indicators across the member states.
- There is an urgent need for a National Health Information System: linking or integrating these existing digital collections of various sources to create a combined or distributed database.
- A unified nationally representative primary care data system that allows the use of data based on the EMR of the Belgian GPs is of eminent importance, and could be established by building on current initiatives.

2. Executive Summary

This report summarizes the work that is performed within the context of the **2019 Pilot Data Collection on diagnosis-based morbidity statistics coordinated by Eurostat**. In this project, we aimed to obtain the best possible, nationally representative diagnosis-based morbidity statistic for each of the proposed indicators from the shortlist as a part of the report on Morbidity Statistics in the EU. Sciensano has seized this opportunity to construct a suitable administrative infrastructure to collect the diagnosis-based morbidity data in a sustainable way, and to facilitate and simplify future work and possible related projects by maintaining the necessary supporting documents and data requests, the code generated for the performed analyses, and the contacts with important stakeholders.

The project required the **involvement of multiple stakeholders** in Belgium who, each in their own way, made unique contributions to the outcomes and results of the tasks and performed work: Intego, the integrated network of General Practitioners in Flanders; The Intermutualistic Agency, that governs and analyses an exhaustive collection of health care and prescription data as processed by the health insurance services in Belgium; and the National Hospital Stay Database, maintained by the Technical cell of the National Institute for Health and Disability Insurance and the Ministry of Health.

The **quality of the data is considered to be variable** due to a number of issues. Based on the data accuracy, completeness and coverage, different quality levels were defined. The validity of some indicators is questioned as these indicators are not considered a medical diagnosis of a specific disease, but rather an indication of a given symptom or syndrome. There are some concerns regarding the compatibility and description of some of the proposed definitions.

Due to the varying quality of the data regarding representativeness at a national level, completeness of the cases and compliance to the Eurostat definitions, we propose an **easy to read scoring system** to evaluate the quality of the different indicators across the member states. This system could be used in the final evaluation of the project to decide on potential changes in the list of indicators.

Finally, to arrive at the best possible, nationally representative diagnosis-based morbidity statistics, there is an urgent need for a **Belgian National Health Information System**: linking or integrating existing digital data collections of various sources. By relying on one source to detect the prevalence or incidence of a given condition, too many cases are potentially missed. At the same time, a lot of time and resources are spent on reaching consensus with all the data holders, adapting the datasets for researchers and building the pipelines for the data exchange. Along these lines, difficulties have been encountered regarding the lack of availability of **nationally representative primary care (diagnosis-based) data**, as there is no system in place which allows to collect such data. A unified nationally representative primary care data system that allows the use of data based on the electronic medical records of the Belgian general practitioners would be very welcome.

3. Introduction

Following the pilot studies on diagnosis-based morbidity statistics, executed in 16 Member States of the European Union (EU) between 2005 and 2011, a series of recommendations on the feasibility of a regular, periodic collection of diagnosis-based morbidity statistics was constructed by the designated Task Force. To examine these findings, Eurostat initiated the 3rd phase of the project and is currently cooperating with 10 EU Member States to identify the existing (national) systems and to evaluate the proposed methods in order to obtain the best possible national estimates for the (diagnosis-based) morbidity indicators, as included in the EU short list as a part of the report on *Morbidity Statistics in the EU*. During this phase, each participating Member State aims to obtain the best possible, nationally representative diagnosis-based morbidity statistic for each of the proposed indicators from the shortlist.

In Belgium, the current phase of the project has been conducted within the Service Lifestyle and chronic diseases at the Department of Epidemiology and public health of Sciensano, the Belgian institute of public health. Sciensano's mission is to provide support for public health policy, national coordination and international representation through scientific research, expert advice and evidence-based reports. Based on independent scientific research, Sciensano formulates evidence-based recommendations and solutions in order to proactively aid to develop national policies regarding public health in Belgium. Sciensano also aims to coordinate health related activities across the country, the communities and regions to facilitate the availability of uniform, qualitative health information to support these recommendations and to aid shaping the health related policies.

The coordination of the current phase of the project was carried out by a senior scientist within the Health Indicators team of the Lifestyle and chronic diseases unit. Additionally, a junior scientist was recruited to perform the data collection and analysis in order to obtain the requested estimates of the diagnosis-based morbidity indicators. Other senior members of the scientific staff within the same unit, as well as the head of the unit, were involved in the project to consult on current and previously performed work and to finalize the files and the report.

In order to obtain the best possible estimates for the diagnosed-based morbidity indicators, Sciensano set out five objectives within the project, where the goal was to:

- **Construct a suitable administrative infrastructure** to collect the diagnosis-based morbidity data in a sustainable way. The infrastructure is intended to facilitate and simplify future work and possible related projects by maintaining the necessary supporting documents, the code generated for the performed analyses, the contacts with important stakeholders and to maintain permissions regarding data that were obtained during the project.
- **Collect the required diagnosed-based morbidity data** according to the complete EU shortlist of indicators within the reference period of 2014-2016.
- **Identify the major restrictions** in obtaining the best estimate for each indicator. By doing so, recommendations regarding the future development of the project and data collections can be

appropriately constructed. Additionally, any glaring flaws in the existing data collection and available data sources in Belgium, as well as possible opportunities that can be exploited, can also be identified, which can greatly improve future data collections and developments in other projects at Sciensano.

- **Develop suitable methodologies** to estimate each indicator, based on the acquired diagnosis-based morbidity data. The development of these methodologies can be considered as an integral part of the aforementioned infrastructure that was set up during the project. These methodologies are intended to simplify future work and related projects and enhance the reproducibility by focusing on the correctness, simplification and documentation of the used methods.
- **Develop potential solutions** to counteract and correct for the identified major restrictions. As the capabilities of the main used sources to generate a best possible national estimate of an indicator, are often limited in one way or another, suitable corrections should be identified and developed to enhance the quality of the output data.

Within the scope of these objectives, each process and required procedure is documented for internal use and to support future work. This report will compile and summarize such documentations in order to create an overview of the progress and current state of the project.

The following report chapters each focus on a specific topic, relevant to the pilot data collection. In **Chapter 4**, the **applied processes and strategies** are clarified, especially regarding the collaborations with the stakeholders of the project in Belgium. The strategy regarding the data source selection is also further clarified here. Subsequently, in **Chapter 5**, the **most important breakthroughs** in the project are indicated, signalling the obtained progress in collecting the best possible, nationally representative diagnosis-based morbidity statistics in Belgium. **Chapter 6** further elaborates on the **difficulties faced** while executing the pilot data collection. The main areas where these complications were found, were the data access, the definitions used by the data sources and the Eurostat guidelines and their (in)compliance and the data quality in general. Then, in **Chapter 7**, some **final thoughts and recommendations** regarding the various successes and difficulties are formulated.

4. Procedures and Methods

4.1. STAKEHOLDERS INVOLVED

The project required the involvement of multiple stakeholders in Belgium who, each in their own way, made unique contributions to the outcomes and results of the tasks and performed work. In **Table 1**, a comprehensive overview of the most relevant organisations, related to this project, can be found, along with their respective roles, indicating their (in)direct influence and how the stakeholder has impacted the activities. Subsequently, the possible contributions and limitations are listed per stakeholder, each indicating the potential strengths (possible contributions) and weaknesses (possible limitations) regarding the involved party. These factors respectively point out the opportunities for each stakeholder, thus indicates where their role is strengthened and how they can actively contribute, and the threats, which indicate to possible damaging factors in the relation between each stakeholder and the projects. The threats are to be considered with great caution as they can substantially limit the validity of the resulting outcomes. Finally, a strategy to maintain the best possible relations between the stakeholders and the project, while making optimal use of the contributing factors and to inhibit the possible limitations of each stakeholder, was developed and incorporated in each activity. A general description of these strategies was then also included in this table.

Table 1. Overview of the stakeholders within the project.

Stakeholder	Role	Possible contributions	Possible limitations	Strategy
Sciensano	Data analysis & Reporting	<ul style="list-style-type: none"> • Deliver diagnosis-based morbidity statistics. • Deliver Belgian Health Interview Survey data (BHIS). • Create internal framework & infrastructure for collection of diagnosis-based morbidity statistics. • Establish new collaborative relations. 	<ul style="list-style-type: none"> • Inability to process the received data into deliverable estimates. 	<ul style="list-style-type: none"> • Weekly meet-ups to discuss encountered difficulties & progress. • Documentation of decisions, actions & analyses.
Intermutualistic Agency (IMA/AIM)	Data Provider	<ul style="list-style-type: none"> • Deliver insurance based data. • Verification of results. 	<ul style="list-style-type: none"> • Delay data access. • Inability to provide requested data. 	<ul style="list-style-type: none"> • Establish project collaboration. • Maintain contact to discuss progress or problems.

Intego Network of General Practitioners	Data Provider	<ul style="list-style-type: none"> • Deliver primary care based data. • Verification of results. 	<ul style="list-style-type: none"> • Delay data access. • Inability to provide requested data. 	<ul style="list-style-type: none"> • Establish project collaboration. • Maintain contact to discuss progress or problems.
National Institute for Health and Disability Insurance (RIZIV-INAMI)	Data Provider	<ul style="list-style-type: none"> • Deliver hospital based data and perform Small Cell Risk Analysis (SCRA) to assess the identification risk. • Provide coding of diseases & health care services. 	<ul style="list-style-type: none"> • Delay data access. • Inability to provide requested data. 	<ul style="list-style-type: none"> • Establish project collaboration. • Maintain contact to discuss progress or problems.
IVC-CSI	Data Protection	<ul style="list-style-type: none"> • Facilitate data request and access permission. 	<ul style="list-style-type: none"> • Not willing to approve data access. • Delay data access. 	<ul style="list-style-type: none"> • Obtain fully justified data requests.
Eurostat	Project Client & Sponsor	<ul style="list-style-type: none"> • Facilitate project. • Facilitate meetings among participating countries. • Provide project grant. 	<ul style="list-style-type: none"> • Cut grant. • Prioritize other projects. 	<ul style="list-style-type: none"> • Participate in meetings. • Maintain contact to discuss updates. • Provide reports.
Federal Public Service on Health, Food Chain Safety & Environment	Government	<ul style="list-style-type: none"> • Facilitation of hospital data in cooperation with RIZIV-INAMI. 	<ul style="list-style-type: none"> • Prioritize other projects. 	<ul style="list-style-type: none"> • Establish project collaboration.

4.2. COLLABORATIONS

Apart from setting up and maintaining the required relations to the relevant stakeholders regarding the project and exchanging contact details, specific collaborations were established to allow the acquisition of data and to facilitate the project. Additionally, some of these collaborations were initiated so that a novel cooperation between Sciensano and the respective third parties was allowed, which leads to enhanced future data exchanges and collaborations within other projects. The experiences while setting up these collaborations and novel data exchanges from this project also proved useful for further use within Sciensano. Specific collaborations with external parties are described below.

4.2.1. Intego Network of General Practitioners

Intego is an integrated network of General Practitioners (GPs) in Flanders, which is the biggest region in Belgium (BE) population-wise and accounts for 60% of the BE population, led by the Academic Centre for General Medicine from KU Leuven, which aims to create a large database to centralise morbidity data in primary care. The network incorporates an automated data collection, based on the Electronic Medical Record (EMR) of the patient as registered in their GP's practice. The collection of such EMR data allows the network to convey the incidence and prevalence of a multitude of (primarily GP, non-specialist care requiring) diseases, as well as data concerning diagnostic tests or applied therapies in the registered GP practices.

The Intego network of GPs was identified to be the main useable source to be able to deliver the diagnosis-based morbidity data in the primary health care in Belgium. Therefore, a collaboration between Intego and Sciensano was initiated to establish the access to the required micro-data in the Intego database. Doing so enabled the determination of the estimates of the GP-based morbidity indicators of the shortlist. Currently, this collaboration is project based and thus will end when the project is finished, however, both parties have indicated to be willing to engage again into more collaborations in the future. It should also be noted that the collaboration refers to a specific scientist at Sciensano due to data protection regulations. Therefore, if the access of multiple scientists is required, multiple collaborative requests have to be initiated in order to be able to establish a full collaboration between all scientists at Sciensano and Intego.

It should be noted however that the data from the Intego Network has several limitations regarding the capabilities to produce a national estimate. Firstly, the geographical limitation, as only GP practices in the region of Flanders are included in the network. Therefore, any regional differences between the Flemish Region, the Brussels Capital Region and the Walloon Region cannot be extrapolated and should thus be determined from a different source (if relevant). Secondly, the network only covers about 2% of the Flemish population. Thus, to produce a national estimate, the prevalence and incidence as calculated from the Intego Network data, should be upscaled in order to obtain the actual prevalent/incident cases in Belgium (while accounting for regional differences when deemed necessary). Thirdly, the diagnoses are coded using the ICPC classification and are, in parallel, mapped using the software thesaurus to ICD-10. However, this mapping is not completely reliable and validated

for accuracy as of yet. Therefore, the ICPC coding system was used to obtain the matching diagnoses for each indicator, even though the ICPC & ICD-10 definitions do not match completely. Also, as the registration in the EMR is not episode based, it can prove tricky to calculate the prevalence for some indicators, depending on the nature and the need to visit the GP during the episode and/or follow-up. It is also unknown how representative the Intego Network of GPs is regarding the GPs in Flanders. Lastly, it is unknown how exhaustive the GP's information on the patient is, when reliance on external sources is desired or necessary (for instance after hospital admission for a specific condition or medical imaging examinations). This may cause some underestimations when the GP is not aware or has no recording of a certain (externally defined) diagnoses. For instance, in Belgium it is not uncommon for a patient to go directly to a dermatologist or gynaecologist, without referral from a GP.

4.2.2. IMA-AIM

The InterMutualistisch Agentschap – Agence InterMutualiste – InterMutualistic Agency (IMA-AIM) governs and analyses an exhaustive collection of health care and prescription data, as processed by the health insurance services in Belgium and aims to support the improvement of the performance, the quality and the accessibility of the Belgian Health Care system and health/invalidity insurance. IMA-AIM also facilitates the Permanent Sample (abbreviated as: Echantillon Permanente Steekproef – EPS), which contains a 1/40 randomly sampled cohort of health insured individuals in Belgium for whom their data regarding health insurance, which is compulsory in BE, is collected over time, allowing to perform observations on an individual level. This cohort is considered to be a representative sample of the whole of the Belgian population (including the three main regions) and was thus identified as the preferred data source to collect the insurance-based morbidity data. The database consists of three types of data, one regarding the population, one containing a database with reimbursed health care procedures and one with the reimbursed medication.

A collaboration between IMA-AIM and Sciensano for the use of the EPS data has been well established in the past and could thus be exploited further, after approval by the technical department of IMA-AIM, to initiate the data access for the scientists at Sciensano that are working on the Morbidity Statistics pilot data collection.

The operational case definition was based on the reimbursed drug prescriptions as a proxy for the disease, corresponding to the definitions as provided by Eurostat. This method however was handled with caution, as using such proxy diagnosis can easily lead to a wrong conclusion. For example, patients not taking the drugs but who do have the disease nonetheless are not identified as a patient when using this approach, which could lead to an underestimation of the indicator. Contrarily, patients taking the drugs for other reasons than the disease of interest (e.g. in the context of drug repositioning) should be excluded as these would lead to an overestimation using such identification approach. To exclude these non-cases, a threshold value to determine a real case is based on the predefined number of Defined Daily Doses (DDDs) per year for each disease.

4.2.3. RIZIV-INAMI

In order to acquire the Hospital Discharge Data (HDD – referred to as Minimale Ziekenhuisgegevens, MZG or Minimal Hospital Data, MHD), the Federal Public Service of Health, Food Chain Safety and Environment (also named the Ministry of Health – MoH) was contacted. The MHD is a registration by the MoH whereby the (anonymised) administrative, medical and nursing data of the (non-psychiatric) hospitals is collected to support the governance and health policies of the MoH and the hospital (networks). However, after initial communication with the MoH, it was clear that the dataset does not allow follow-up of patients over multiple years, which makes it impossible to distinguish incident cases from prevalent cases, and between hospitals. The latter can prove particularly tricky, as in Belgium, the patients have the right to choose their own caregiver and are thus not bound to go to 1 particular hospital. This would lead to a misclassification of an incident case when a patient would, for example, visit a second hospital for follow-up, rather than continuing his/her care in the same hospital. To solve this problem, it was proposed to use the National Hospital Stay Database (NHSD) rather than the MHD.

The NHSD is a merged dataset, which is based on:

- The MHD from the MoH, which contains information concerning the diagnoses and procedures for each admission.
- The Hospital Billing Data (HBD) from the National Health Insurance companies, which contains the information regarding the billing data for hospitalized patients. This data was sent by the hospitals to the health insurance companies for reimbursement.

The primary goal of this dataset, which is created by the TCT (Technische Cel – Cellule Technique) from RIZIV/INAMI, is to generate a comprehensive overview on the required care, the financial needs and the reimbursed costs, for each specific pathology and to be able to list these costs for each hospital individually and compare the data between all Belgian hospitals. Additionally, by combining the MHD and the HBD, it is possible to create a linkage between the patient and the received care over multiple years and between multiple hospitals. Therefore, the NHSD was preferred over the MZG to deliver the required information on the HDD.

Since there have been no previous official data exchanges of the NHSD yet, a new collaboration between the TCT of RIZIV and Sciensano was set up to enable access to the required data. The use of the NHSD database is considered to be advantageous for both current and future projects within Sciensano as this database contains valuable epidemiological data, which can be used to strengthen findings and evidence based policies. Both parties expressed their interest and enthusiasm regarding such cooperation, where the TCT would deliver the data if Sciensano would perform the data analysis. However, as this collaboration was novel and requires the necessary permissions from the governing committees regarding data privacy, acquiring these permissions to set up the data exchange proved to be a difficult and cumbersome process. It should be noted that, with the necessary permissions acquired, not only is the data exchange within the scope of this pilot data collection permitted, but future data exchanges should be permitted as well. Once the initial permissions are acquired, obtaining future permissions should prove less cumbersome, which will simplify routine data requests.

4.3. DATA SOURCE SELECTION

The data source selection builds further on the findings from the two previous phases in the Morbidity Statistics project while accounting for the definitions and requests within the current phase of the project.

It should be noted however, that no additional new (useable) data sources were identified at the start of this phase when compared to the previous phases, while some data sources that were included in the previously proposed list of data sources were deemed to be not appropriate within the framework of the specific objectives and current definitions.

The data sources, as used in the pilot data collection, can essentially be divided into four categories of morbidity data, based on their basic composition and origin, allowing to cover most of the required health care areas to account for the requested diagnosis-based morbidity data:

- Specific registers
- Primary care data
- Hospital discharge data
- Health insurance data

Unfortunately, as was also concluded in the second phase of the Morbidity Statistics project, there is still no system in place to routinely link these types of data sources at an individual level in Belgium. As some kind of workaround, the aggregated data from e.g. the Cause of Death Registry can be added to the Hospital discharge data to account for deaths before arrival in the hospital, but routine linkages between sources is not yet available. Through the morbidity statistics project, multiple discussions among the various stakeholders have been held to start developing such initiative. However, even though most parties agree on the added scientific value, most are hesitant regarding the practical implementation. Reasons for this are mainly related to financial doubts and privacy regulations, because such a system could easily become expensive to implement and maintain, while the validity regarding the current privacy regulations is doubtful. To get the permissions to initiate and operate such a system will most likely prove difficult without the creation of a legal framework. Therefore, new in-house meetings have been set up to start thinking about the practical implementation and the necessities of to start up and maintain such a system. Also, by developing a more global vision, the benefits for all involved stakeholders can be clearly set out. These discussions should lead to a more constructive initiation of the system and a possible practical roadmap to develop a national health information system.

4.3.1. Selection of the best sources

To obtain the final data sources for the pilot data collection, first, the information and conclusions as proposed in the second phase of the Morbidity Statistics project were used and evaluated to check if these data sources are still compliant to the proposed definitions. To do so, the concept notes, where the data source is evaluated in function of each disease, from the second phase of the project were used. The selection of the sources was as follows:

1. If the proposed data source from the second phase was still compliant to current definitions and considered to be the best available source to provide the best possible national estimate of the indicator, then the data source is kept as the primary data source for the pilot data collection.
2. If the proposed data source from the second phase is unable to comply to the current definitions or if the proposed data source is unable to produce the best possible estimate of the indicator, then the data source is replaced by an alternative source for the pilot data collection.
3. If there was no proposed data source, a data source was proposed to be able to produce a national estimate for the indicator.

The evaluation and proposal of alternative/new data sources was conducted similarly to the evaluation in the second phase of the Morbidity Statistics project using the following criteria:

4. Is the source based on medical diagnoses or a proxy diagnoses (e.g. the use of prescription medication)?
5. Is the source exhaustive or sample-based?
6. If the source is sample-based, is the sample population well defined and representative?
7. Does the source capture all cases? Or in the case of a sample, is the source exhaustive for the covered population?
8. Is the coverage regional or national?
9. If the source is regional, how can the regional differences be accounted for (if existing)?
10. Does the source allow the computation of annual figures?

The complete list containing the indicators and their respective sources, as used in the pilot data collection, can be found in **Annex A**.

5. Development and Successes

5.1. ACQUIRED DATA

To acquire the necessary diagnosis-based morbidity data, new collaborations were set up and existing partnerships were further exploited. The acquired data is individual-based and was pseudonymised by the data source organisation or a Thrusted Third Party (TTP) facilitating the data access. This allows to precisely calculate the desired indicators following the proposed definitions, rather than making use of (previously reported) aggregated data. The four main diagnosis-based data types, as mentioned in Section 4.3 Data Source Selection, are discussed below.

5.1.1. Primary care data

As discussed in Section 4.2.1, the main data source used for the primary care data is the Intego Network. This data is based on EMR registrations and thus includes the following important information:

1. Basic patient attributes
 - Gender
 - Date of birth
 - District of GP
 - Yearly Contact Group¹ (YCG)
2. Diagnosis
 - Start of Diagnosis
 - End of Diagnosis
 - ICPC Code
 - ICD-10 Code

The patient attributional variables allow to calculate the indicators, broken down by gender and age (which is calculated using the date of birth). It is not possible to break down the data between residence status as only people with a social security number are registered (and are thus by definition only residents). The diagnosis variables then allow to specify the diagnosis by ICPC-2 classification, as the software is based on ICPC classification coding of the diagnoses and the mapping from the software thesaurus to ICD-10 has proven to be suboptimal. This does cause some compatibility issues regarding the definitions as proposed in the shortlist of indicators, however, the used ICPC codes were the codes as proposed in the shortlist as an alternative for the ICD-10 definition of the indicators. The start of the diagnosis is essential to determine if a case is either an incident case or not, while the end of the diagnosis can be used to exclude cases with past episodes (for a non-chronic disease) when calculating the prevalence. The calculation of the incidence by person and incidence by episode is thus rather straightforward, as the exact onset of the diagnosis is included in the record.

¹ The YCG is the number of patients who visited their GP during the year of reference

The calculation of the prevalence, however, is not as straightforward, as the registration in the EMR does not record the primary incentive for each visit, which thus makes it impossible to clarify whether a visit to the GP is linked to the previously diagnosed disease or not. Therefore, the calculation of the prevalence should be based on the Yearly Contact Group (YCG) or the Practice Population (PP). Although, in Belgium, the patient is free to choose their GP, often, it is opted that a patient can register their Global Medical Dossier (GMD) with a GP of their choosing. Usually, the patient will then visit this GP most regularly as then, the GP will have the best (medical) knowledge of the patient. The YCG translates then to the patients who contacted their GP at least once in that year. Therefore, all patients who contacted their GP in 2016 will be indicated as YCG = 2016. There are however patients who do not visit their GP during that year. To account for these non-visits, the Practice Population (PP) can be calculated. The PP is the estimated total of patients of each GP practice, whether they visited their GP during that year or not and is calculated using the YCG and a correction factor to account for the fraction of non-visiting patients (corrected for gender, age and district). The PP was preferred over the YCG for the calculation of the prevalence as the PP was deemed more representative of the Belgian population. Using the PP, the prevalence of the indicators could eventually be calculated by counting all cases from the PP who visited the GP in the reference period who were once diagnosed with the disease. However, the results based on this calculation should be treated with caution as not every indicator is suitable to use this method to calculate the prevalence.

5.1.2. Insurance-based prescription data

The EPS from IMA contains information as registered by the health insurance services in Belgium and thus allows to calculate the indicators based on reimbursed insurance-based (prescription) data. An extensive use of prescribed medication intake can be used, using the ATC coded medication as a proxy for the disease. There were basically two types of definitions of the diagnosis possible: A predefined pseudo-diagnosis as created by an expert group, appointed by IMA/AIM and RIZIV/INAMI; or a custom created pseudo-diagnosis.

By default, the EPS provides predefined pseudo-diagnoses, defined by a working group set up at RIZIV/INAMI, as a collection of ATC codes with a DDD threshold, which can be used directly from their EPS database. By doing so, the incidence and prevalence can be calculated easily. Should the case definition not match any predefined pseudo-diagnoses or should there be no pseudo-diagnoses defined, users can create their own case definitions using the ATC codes and DDDs using a sub-file (Pharmanet) of the database. In either case, the proxy for the diagnosis can be calculated using the provided ATC codes from the proposed definition. Additionally, the data can be broken down by gender, age (calculated from birth year) and official residence, which is determined by the domicile region of the person and is thus not an exact match of the residence definition as proposed in the guidelines.

The prevalence was then determined by counting the number of people corresponding to the DDD threshold as set in the EPS within the reference period. The incidence (by person) was then estimated by counting only the cases which had such medication use in the reference year, but not in the preceding two years of the reference period.

5.1.3. Hospital inpatients

The National Hospital Stay Database (NHSD) is a merged database, linking financial (information regarding the billing data for hospitalized patients) and clinical (information on the diagnoses and procedures for each admission) aspects for each stay in a Belgian hospital, thereby only including hospital inpatients, not outpatients. Permission to obtain access to the National Hospital Stay Database (NHSD) was granted after a Small Cell Risk Analysis (SCRA) to assess the identification risk. For the current project, pseudonymised patient number; year of registration; registration period; year of admission; month of admission; year of discharge; day in the month of discharge from the hospital; gender; district; verified admission diagnosis; year of birth and main/secondary diagnosis were available for 2014 and 2016.

As data for 2015 was not available in the NHSD, for the calculation of the incidence by person and incidence by episode the episodes that began before the reference period (2016) could not be excluded. All episodes that began during the reference period were included. The record of a hospital inpatient episode is likely to contain not only the main reason for admission but mention of other health conditions present at the time of admission, and also symptoms and complications. To be counted as a prevalent case, it was not necessary for a condition to be the main reason for admission; secondary diagnoses were also included.

5.1.4. Survey data

The Belgian Health Interview Survey (BHIS) is the health survey in Belgium carried out every 4 to 5 years. Each wave comprises of about 10,000 surveyed individuals in some 6,000 households. The purpose of the BHIS is to monitor the health status of the general population as well as health determinants including health behaviours, medical care consumption and social and demographic characteristics. The target population of the BHIS consists of all persons with residence in Belgium, including the institutionalized elderly, with no restrictions on age or nationality. The National Population Registry (NPR) is used as the sampling frame. This registry contains information on gender, age, address, citizenship, marital status, etc. of each individual. Data collection is undertaken using face-to-face interviews at the participant's home. Results of this survey can be accessed using HISIA: a web-based interactive tool that presents the weighted percentages or means of selected variables of the BHIS.

As the Intego network contains only the information of GP practices in Flanders, who voluntarily joined the network, only 2% of the population within the region of Flanders is covered, which can hardly be considered as a representative sample for the Belgian population due to the regional differences. Using regional corrections, based on the weighted percentages from HISIA, that can be regarded as nationally representative source, to estimate representative numbers for diagnostics based on Intego data.

5.2. ANALYSIS INFRASTRUCTURE

One of the key objectives of the project is to set up the required infrastructure to support the data access, collection, analysis and organisation, while facilitating future work.

5.2.1. Used software

The standard office tools were used for basic organisational tasks like generating reports, spreadsheets, documentation files, planning, ... The files that were used, were documented and created in such a way that future use of the tools is as easy as possible.

For the collection and analysis of the data, the main used software was SAS and R. The access to some of the data sources is facilitated by a TTP, which requires an active licence of SAS. Therefore, to extract the data from these sources, reusable and well-documented SAS programs were written. Once the data are obtained, the analysis was performed using R. Again, the used code was written to be as easy to read as possible with a focus on ease of use and reproducibility of the results, so that, no matter who would use the program, would obtain the same results without having to interfere too much in the code. To do so, the R *shiny* package was used. Therefore, to analyse the data, the researcher can use a shiny web application to explore the data, call a visualisation and generate the output for the selected indicators, without having to change the code. For the analysis, the *mgcv* package was used to generate a Generalised Additive Model (GAM) of the prevalence/incidence data. Using such a model allows to approach the binary state of an individual (case vs non-case) through a binomial probability distribution and smooth the data, while accounting for the given variables as gender, residence, age.

5.2.2. Analysis methods

To increase homogeneity of the results and to streamline the analysis, the same approach was used to analyse the data for all data sources. This approach consists of three steps:

Select the relevant data in the dataset:

- Only the data regarding the reference period was selected from the database to exclude non-relevant data.
- The cases of interest were selected using the classification system of the data source (ATC, ICPC, ICD-10) and the corresponding criteria.

Calculate the indicator

- A target variable (y) based on the definition of the prevalence or incidence was created, where the following is true:
 - Target variable for prevalence: $y = 1$ if the patient is considered a case in either of the years within the reference period. Otherwise: $y = 0$.
 - Target variable for incidence: $y = 1$ if the patient is considered a case only in the reference year. Otherwise: $y = 0$.
 - **Table 2** can be considered to aid the understanding of the target variable.

- To account for extreme deviations in the data, an estimation GAM model based on a function in the form of: $y \sim \text{Age} + \text{Gender} + \text{Residence}$, was generated. This model provides data smoothing and is able to break down the data to the desired category, with respect to the total population.

Upscale the calculated indicators to the Belgian population

- The total number of cases in Belgium was calculated using the mid-year population data of the reference year (2016) as the reference population.
- The upscaling of the estimated indicator was performed by each corresponding category, broken down by age, gender and residence.

Table 2. Truth table containing the definitions of when the case can be considered as a prevalent case or incident case, based on the case status in the reference period [T-2, T] with T = 2016.

Case in 2014	Case in 2015	Case in 2016	Prevalent case	Incident case
0	0	0	0	0
0	0	1	1	1
0	1	0	1	0
0	1	1	1	0
1	0	0	1	0
1	0	1	1	0
1	1	0	1	0
1	1	1	1	0

5.2.3. Tools & software development

As mentioned before, some tools were developed to increase the reproducibility of the analysis and to decrease the risk of errors due to miscoding or the collection of variables. First, to facilitate the data collection, which required in most cases the use of SAS, reusable SAS macros were written to ensure a homogenous information collection among the indicators and future extractions. By doing so, the researchers working on the data collection have a clear view on the necessary variables, which are required to perform the analysis, and can be certain that the data collection includes the same variables each time it is performed.

To perform the analysis, it was opted to develop an application using R to assist the researcher to analyse the data and to enable the user to explore the data without having to code. By doing so, and effectively separating the user & development aspects of the analysis, the risk of coding errors due to necessary changes is reduced. This application is intended for in-house use only, with only the researchers working on the project being able to access the app. Therefore, it was possible to include the data exploration aspect as this function shows the actual microdata from the source. A screenshot of the application in use can be found in **Figure 1**, which shows the interface and the possible graphical

output based on the data and the model. Using the interface, the researcher can indicate whether or not to show the source or model data (the latter including the variables used in the model and the target variable), which indicator is used (prevalence/incidence) and for which disease the indicator is calculated. When this selection is performed, a graphical representation of the source data and the indicator estimation can be shown (as shown in **Figure 1**) and the output data, which matches the guidelines format, can be generated.

Along with the comments and descriptions, provided in the code, tutorials and user manuals were crafted to aid future iterations of the project and related assignments using similar analysis techniques and data. This will provide especially useful in communicating the applied analysis and data extraction techniques as most of the data that was received was obtained in a novel way and had to be explored before exploiting these datasets. During future iterations and related projects, when data should be extracted from the same sources, these tutorials can be used to shorten the time, necessary to apply for access to the data and the data exploration. These tools are an integral part of the infrastructure that was set up to complete the pilot data collection.

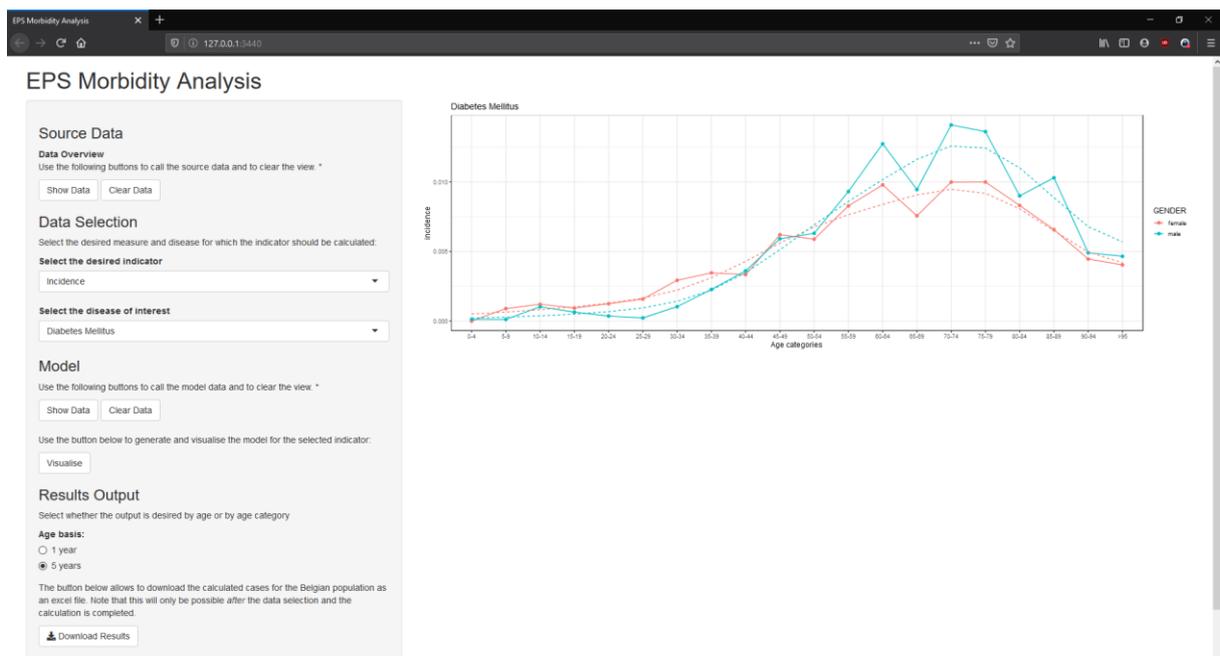


Figure 1. Screenshot of the application used to analyse the EPS data. On the left, multiple fields can be recognised where the user can indicate the desired actions (e.g. show the source data, model the data, create a visual representation and generate the output).

5.3. CORRECTIONS

As mentioned before, the data from the Intego network of GP practices is limited to Flanders and, as was shown before, there are significant differences in the prevalence of certain diseases among the different regions in Belgium. Therefore, when relying on data, based on only one of these regions, corrections should be applied to calculate a representative national estimate. The results, based on non-national data, can be corrected by taking into account the prevalence/incidence ratios between the regions from HISIA (self-reported data source, see Section 5.1.4). **Table 3** shows the indicators and the correction factors from BHIS that were used to calculate a representative national estimate. The results from BHIS 2018 were taken, as this was closest to the reference year of 2016.

Table 3. Indicators from the Intego network and the correction factors derived from the Belgian Health Interview Survey.

Indicator	Item in Belgian Health Interview Survey
P4 — Mental & behavioural disorders due to alcohol use	Percentage of the population aged 15 years and over ever experiencing problematic drinking (CAGE, 2+)
P6 — Mood (affective) disorders	Percentage of the population aged 15 years and over with any type of depressive disorder (PHQ-9 criteria)
P7 — Anxiety disorders	Percentage of the population aged 15 years and over with a generalised anxiety disorder (GAD score 10+)
P11/P12 — Hypertensive diseases	Percentage of the population aged 15 years and over that suffered from high blood pressure (hypertension) in the past 12 months
P13 — Ischaemic heart diseases	Percentage of the population aged 15 years and over that suffered from coronary heart disease (angina pectoris) in the past 12 months
P16 — Heart failure	Percentage of the population aged 15 years and over that suffered from myocardial infarction in the past 12 months
P24 — Alcohol liver disease	Percentage of the population aged 15 years and over ever experiencing problematic drinking (CAGE, 2+)
P25 — Diseases of liver other than alcoholic	Percentage of the population aged 15 years and over that suffered from cirrhosis of the liver, liver dysfunction in the past 12 months
P26 — Diseases of liver	Percentage of the population aged 15 years and over that suffered from cirrhosis of the liver, liver dysfunction in the past 12 months
P28 — Arthrosis	Percentage of the population aged 15 years and over that suffered from osteoarthritis (arthrosis, joint degeneration) in the past 12 months
P29 — Osteoporosis	Percentage of the population aged 15 years and over that suffered from osteoporosis in the past 12 months

6. Encountered Difficulties

6.1. ACQUIRING DATA ACCESS

Due to privacy regulations, acquiring access to the microdata from various sources has proven difficult and a cumbersome process. The reason being that, within the context of specific research projects, the use of pseudonymised individual (personal) data can only be authorised after a complex procedure of investigating the security risks of each dataset and having received positive advice from the Information Security Committee. Additionally, to comply to the privacy regulations, the patients should be aware of the information that is being processed (by the means of a clause or personal contact, of which the latter would not be possible in this case), or a legal framework should allow such processing. The latter is applicable to the EPS, where access is granted structurally after initial request by the Sciensano researchers.

The confidentiality of the data was considered with great care. For most sources, a remote access system was used to grant access to the data, rather than making the (encrypted) dataset directly available for analysis. By doing so, data leaks are supposedly avoided. However, as each data source uses their own (proprietary) system, creating a linkage between the data from one source to another is not possible. Therefore, organising a linkage between the datasets is, for the time being, not considered to be feasible, unless a legal framework is developed and such linkage can be granted by law.

Regardless of the great care regarding the privacy protection and the confidentiality measures that are in place, the access was severely limited and permissions were obtained with considerable delay. Consequently, efforts were made to ensure a smooth cooperation between Sciensano and the data sources in order to facilitate routine data collections in the future without having to reiterate the data requests yearly.

6.2. DEFINITION AND DATA COMPLIANCE

Difficulties were also encountered regarding the proposed definitions in the guidelines and the possibilities within the available data. There were basically two types of discrepancies possible, regarding the compliance towards either the indicators or the population, as discussed below.

6.2.1. Indicator definitions and code mappings

The case definitions as proposed for the indicators are based on the ICD-10 classification system, however this classification is not always available in the used sources.

The Intego GP Network uses the ICPC classification, where the GP can indicate a disease case, based on a selected set of definitions in a thesaurus for each diagnosis. Therefore, the underlying concept

can be different or a slight mismatch to the definitions as used in the ICD classification. The proposed ICPC codes from the shortlist were used where possible.

The EPS dataset is based on the proxy definition of the diagnosis, characterised by the ATC classification. Where possible, the proposed ATC definition from the shortlist is used. However, by using this classification, some mismatch in the estimation might arise as there might be identifiable cases not using the medication for the listed ICD-10 defined disease due to e.g. drug repositioning. An example of this might be obese patients using diabetes related medication (categorised as ATC A10) for the purpose of weight control.

The NHSD used the ICD-9 classification prior to 2015 and has adopted the ICD-10 classification since. Caution is advised when interpreting the results prior to and after the transition between the classification systems as the use of the different classifications might yield a difference in results due to mismatching between the systems. Where applicable, the ICD-9 proposed definitions were used for the data prior to 2015 and the ICD-10 proposed definitions were applied to the data after the transition.

6.2.2. Population definitions and available data

Other discrepancies were found in the definitions regarding the case populations and the respective available information in the datasets. More specifically, the exact definition of residence and thus the distinction between residents and non-residents, as was proposed in the guidelines could not be applied within the context of the used data sources.

The Intego GP network dataset only registers patients with a Belgian Social Security Number and only indicates the district of the GP practice where these patients are registered. Therefore, no information is available on non-residents. The estimated population of non-residents, treated by the GPs attached to the network, is considered to be less than 1% of the total treated population. Regional coverage of the Intego data source is limited to Flanders only. Furthermore, the use of Intego data can lead to false negatives; the bypass of the GP when consulting direct specialised care will lead to an underestimation of the actual prevalence.

In the EPS dataset, the residence of the patient is determined based on where the patient has his/her domicile registered. It is therefore possible to see regional differences within Belgium for a certain disease, which would enable the EPS data to serve as a regional correction if the EPS was not considered as the primary source. Additionally, non-residents can be identified by having a foreign domicile. This is however not completely compliant to the proposed definition as these non-residents could include people living in Belgium for more than 12 months, but still having their official address in a foreign country. On the other hand, patients, with their official address in Belgium, living for more than 12 months abroad, who return for medical care to Belgium, would then be seen as residents. This is however the only possibility to include the residence of the patient into the statistics for this dataset.

In the NHSD, the residence of the patient is determined based on where the patient has his/her domicile registered. Based on the postal code this was aggregated into regional levels, thereby including all individuals who set up residence in a Belgian municipality.

6.3. DATA QUALITY

The quality of the data is considered to be variable due to a number of issues. These issues can be divided into three main categories concerning the accuracy, the completeness and the coverage of the data, which are defined in Sections 6.3.1 to 6.3.3. In Annex A, where the overview of each indicator and its preferred source is listed, a quality indication is also included. Here, one can recognise five quality levels, grouped into three colours:

- A) **Good quality**, which means the data source is considered to deliver qualitative diagnosis-based morbidity data. Indicated in green in the table.
- B) **Good coverage & completeness**, which indicates good data quality regarding the coverage and completeness of the data while some issues are present regarding the definitions and/or methods used. Also indicated in green in the table.
- C) **Doubtful coverage, completeness & methods**. There are issues identified regarding the completeness of the cases and coverage of the population in the data, in addition to the issues named in quality level B. Improvements are deemed necessary to be able to generate a representative national estimate. Indicated in yellow in the table.
- D) **Doubtful quality**. There are diagnostic data available, however, the data quality is evaluated so that the data are considered to be very divergent from the Eurostat definition. Other sources should be considered to report data on this topic. Indicated in red in the table.
- E) **No data available**. No diagnostic data could be provided for these indicators. Indicated in grey in the table.

These quality levels try to comprise the different aspects of the quality categories as discussed below to indicate whether the data can be used for official statistics or whether there is room for improvement.

6.3.1. Coverage of the population

The first important qualitative aspect of the data which can be evaluated, is the coverage of the data, in order to inspect how well the population is represented in the dataset.

The NHSD is a national database, which thus includes all hospitalised cases in Belgium and is regarded to be quasi-exhaustive for the residential population.

The EPS database is a random sample of 1/40th of the Belgian insured population, which is 99% of the residents. This data source can thus be regarded as a national representative of the Belgian population. The non-resident population is also covered in this database, as they are indicated using a unique code,

which makes it possible to account for non-residents. As the database is based on a sample, the total number of cases has to be calculated by multiplying the estimated prevalence/incidence with the total population size. The obtained results for the indicators are therefore estimations.

The Intego network only contains the information of GP practices in Flanders, who voluntarily joined the network. Therefore, merely 2% of the population within the region of Flanders is covered, which can hardly be considered as a representative sample for the Belgian population due to the regional differences. However, using regional corrections, based on nationally representative sources like the BHIS, the limited coverage can be accounted for. Furthermore, non-residents are not registered, as the data collection is based on the EMR of the patient, which needs a Belgian social security number for identification. Analogously to the EPS database, due to the limited coverage, the total number of cases in Belgium needs to be calculated and is therefore also an estimation.

6.3.2. Completeness of the cases

The second qualitative aspect of the data considered, was the completeness of the cases regarding the indicators. None of the sources are considered to be exhaustive regarding all of the cases of a given disease, with the exception of a few indicators requiring specific conditions.

The NHSD is a national database, which includes data from all hospitals in Belgium and can thus be considered to be exhaustive for hospitalized patients. However, cases that do not require hospitalisation are missed, which would lead to a severe underestimation when using the NSHD for diagnoses that do not always require admission to hospital.

The EPS database contains information regarding medication use as a proxy for the diagnosis, rather than the actual diagnosis itself. However, the relation between the medication and diagnosis is often not sensitive enough to include all diagnosed cases (e.g. patients not taking the medication for treatment, using alternative therapies,...), thus generating false negatives. Ideally, the specificity should be so that patients, who were not diagnosed with the disease of interest, but are taking the proposed proxy medication, are not included, thereby avoiding false positives.

Regarding the Intego network, in Belgium, the GP has no gate-keeper function. Therefore, patients who directly seek specialist care could be missed if not properly communicated to the GP. The information could also be missing if the external information is not stored or recorded by the GP within the EMR. Additionally, the completeness of the data depends on how the EMR is used by the GP, as the collected information is based on routine data, and it is unknown to what extent and how detailed all diagnoses are recorded.

6.3.3. Accuracy

The third and last aspect of the data quality considered was the accuracy or the correctness of the data regarding the information obtained from the indicators.

The accuracy of the NHSD data might be affected by the primary purpose of the data collection, which was financial rather than epidemiological, as the diagnoses are registered as part of the administrative system of the hospitals.

The EPS database, being based on health insurance data, is also based on administrative data and could thereby suffer from the same accuracy problem as the NHSD. Additionally, fraudulent use of preferential reimbursements can deter the accuracy as, in this case, the person whom received the prescription is not the person who actually takes the medication.

The accuracy of the data from the Intego network database is relatively unknown and depends on a number of factors. First, the data are based on routine data, registered by the GP with the clinical purpose of case management. Consequently, the quality of these data may vary by GP and could be insufficient due to lack of time or interest. As mentioned before, it is also unknown to what extent external diagnoses (e.g. from specialist care or lab results) are recorded by the GP. Lastly, as the data is case management-based, it is not always possible to identify a visit to the GP as a follow-up visit for a given condition or for a different disease.

6.3.4. Set of indicators

Some concerns regarding the set of indicators were raised during the execution of the pilot data collection. Many of these issues can be traced back to the definitions linked to these indicators and will be discussed in the next section. The added value of some proposed indicators is questioned, especially in the context of collecting diagnosis-based morbidity data. The validity of some indicators in the proposed list is questioned as these indicators are not considered a medical diagnosis, but rather a cause of a given condition. These indicators are:

6.3.4.1. P8 — Parkinson's disease

The results, obtained by analysing the prevalence/incidence of Parkinson's disease, should be analysed with caution as a difference exists between Parkinsonism and Parkinson's disease. Treatments exist to reduce the Parkinson-like symptoms that are not part of Parkinson's disease, but are initiated due to a different mechanism of action. These Parkinson-like symptoms are defined as Parkinsonism and can be mistakenly diagnosed as Parkinson's disease, especially using proxy diagnoses.

6.3.4.2. P20/P21 — Asthma & P23 — Chronic obstructive pulmonary disease (COPD)

The differentiation between asthma and COPD is not always straightforward, even for the physician diagnosing the patient. Therefore, even when using diagnosis-based morbidity data, caution should be exercised when interpreting these results. Understanding the evolution of these important chronic diseases, however, is tremendously important to initiate adept policies regarding the treatment.

6.3.4.3. PB36/PB37 — Land transport accidents & PB38/PB39 — Accidental falls

Although there are ICD-10 codes to indicate a land transport accident, due to the nature of the data (which are administrative, primarily focussed on the financial aspects of secondary care), a primary focus on the given care might exist. This leads to the inclusion of the codes regarding fractured bones or other traumas rather than indicating an accident took place. Further examination of the dataset is needed to examine these concerns. For accidental falls, similar concerns are expressed, as a fall might cause a diagnosis of trauma, which would be most likely to be included in the diagnosis, rather than the fall itself.

6.3.4.4. PB40/PB41 — Intentional self-harm (incl. suicidal attempt)

The indicator could experience the same problems as the aforementioned indicators in this list. Additionally, the definition of the indicator is rather broad, as the indicator includes anything between suicidal attempts, intentionally inflicting cuts or intentional self-poisoning. This could lead to a significantly varying case definition depending on the sources available to investigate the indicator, leading to vastly different results among the member states. This would render the indicator virtually unusable for (inter)national comparisons.

6.3.4.5. PB42/PB43 — Complications of medical and surgical care

The same concerns regarding the broad spectrum of possible case definitions could be made, as the indicator could include anything between a minimal allergic reaction to antibiotics to major complications during surgery leading to the death of the patient. The ambiguity found within the indicator leads to a potentially unusable indicator.

6.3.5. Definitions used

As discussed before, the data is unable to comply to some of the proposed definitions. Specific issues were mentioned regarding the proposed definitions of residence. Some of the consulted sources determine the residence of the patient based on their domicile address, not on their length of stay in Belgium. Some sources even have no case definition for non-residents and are thus not capable of registering non-residents by default. As a result of this, regardless of the fact that it would be possible to register residence duration using administrative data sources, such as health insurance data, the resulting indicators are limited to the available case definitions and the residence is divergent from the proposed case definition. Therefore, the added value of the continuous use of the currently proposed definition of residence is questioned.

Most of the consulted data sources also rely on different disease classification systems, such as ATC and ICPC-2, therefore also creating a (slight) diversion from the case definitions. However, where possible, the proposed alternative disease classification codes by Eurostat were used. Consequently, the results still comply to the, albeit secondary, proposed definition; however, concerns remain regarding the compatibility and description of some of the proposed definitions:

6.3.5.1. P19 — Pneumonia

The ICPC-2 code (R81) is not completely matched to the original proposed ICD-10 definition (J12-J18). In addition, this indicator could not be corrected by using the BHIS, as no item on Pneumonia was available. An alternative source for correction could be sought in other databases, such as the NHSD or cause of death register data. However, here only the most severe cases will be included.

6.3.5.2. P22/P23 — Chronic lower respiratory diseases other than asthma/Chronic obstructive pulmonary disease (COPD)

Identical ICPC-2 codes are used. Therefore, identical results will be produced based on this definition, while the original ICD-10 proposed definition would imply (slightly) dissimilar results. Merging the two disease into one category would provide a solution.

6.3.5.3. P24/P25 — Alcohol liver disease/ Diseases of liver other than alcoholic

As discussed in the previous phase of the Morbidity Statistics project, there is no possibility to differentiate the alcoholic and non-alcoholic liver diseases in primary care (both are D97 coded). Therefore, only the pooled results can be precisely determined. Measures to estimate the proportion of alcoholic and non-alcoholic based liver diseases are being thought of, but no guarantee exists these estimations will be able to generate precise estimations. In an effort to differentiate between the two diseases, P24 was corrected by a correction factor based on the BHIS item on (self-reported) problematic drinking, whereas P25 was corrected by a correction factor of the BHIS based on (self-) reported cirrhosis of the liver and liver dysfunction. Despite their pathophysiology being different, merging both P24 and P25 indicators could provide a solution. When using these statistics for, for instance, measuring impact on health care needs, a merged P24/P25 should be sufficient. However, when focusing on prevention, the difference between P24 and P25 is important to take into account.

7. Recommendations

Key points:

- We propose a **reduced set of indicators** for inclusion, as the validity of some indicators is questioned and there are some concerns regarding the compatibility and description of some of the proposed definitions
- We propose an easy to read **scoring system** to evaluate the quality and comparability of the different indicators across the member states
- There is an urgent need for a **National Health Information System**: linking or integrating these existing digital collections of various sources to create a combined or distributed database
- A unified nationally representative **primary care data system** that allows the use of data based on the EMR of the Belgian GPs is of eminent importance, and could be established by building on current initiatives.

7.1. INDICATOR SHORTLIST

Based on the issues encountered during the pilot data collection phase, we propose the following modifications to the Eurostat shortlist of indicators for diagnosis-based morbidity statistics.

Table 4. Proposed modifications to indicator shortlist.

Indicator	Disease	Quality	Recommendation
P19	Pneumonia	Doubtful	Examination of ICPC and ICD in order to reach consensus on their matching
P22/P23	Chronic lower respiratory diseases other than asthma/Chronic obstructive pulmonary disease (COPD)	Doubtful coverage, completeness & methods	Merge the two indicators
P24/P25	Alcohol liver disease/ Diseases of liver other than alcoholic	Doubtful	Merge the two indicators, depending on the intended use
PB36/PB37	Land transport accidents	Doubtful	Drop from shortlist
PB38/PB39	Accidental falls	Doubtful	Drop from shortlist
PB40/PB41	Intentional self-harm (incl. suicidal attempt)	Doubtful	Drop from shortlist
PB42/PB43	Complications of medical and surgical care	Doubtful	Drop from shortlist

7.2. DATA QUALITY EVALUATION SYSTEM

Due to the varying quality of the data regarding representativeness at a national level, completeness of the cases, compliance to the Eurostat definitions, ... a comprehensive, yet concise quality evaluation method should be applied. This creates a clear overview of the quality of the data Belgium is able to deliver to Eurostat. Additionally, as a lot of heterogeneity is expected in the quality of the data as delivered by the different member states, it will be useful to use an easy to read scoring system to evaluate the quality and comparability of the different indicators across the member states.

We propose to use a colour-based evaluation system as used in Belgium to evaluate the data quality, which consists of:

- **Green:** The data are of good quality with respect to the three criteria — good completeness, good coverage of the population and good accuracy. Or the data has a good coverage of the population and a good completeness of the cases and is thus nationally representative, while some issues are identified regarding the methods and/or case definitions. e.g. the residence in the data source is defined differently than as was proposed in the Eurostat guidelines.
- **Yellow:** The quality of the data can be considered as doubtful. There are issues identified regarding either the coverage of the population or the completeness of the case, thereby making the source, without corrections, not nationally representative. Alternatively, if large issues regarding the methods and/or definitions are identified, this level is applicable as well.
- **Red:** The data quality is considered to be poor. The definitions and/ methods used by the data source are too divergent to be compliant to the Eurostat proposed indicators. Other data sources should be proposed to deliver a reliable nationally representative estimate.
- **Grey:** No diagnosis-based morbidity data are available on the indicator.

Using these four colour codes creates an easy to read, yet clearly defined, quality evaluation, which can be used to easily compare the delivered data in a summarising table. Such comparison could be made as such:

Country	Indicator 1	Indicator 2	Indicator 3		...	Indicator <i>i</i>
Country A	Green	Yellow	Green	Red		Green
Country B	Yellow	Green	Green	Grey		Red
Country C	Red	Green	Yellow	Green		Green
Country D	Yellow	Green	Grey	Green		Yellow
...						
Country Z	Red	Green	Red	Green		Green

This system could be used in the end evaluation of the project to assess the necessary improvements for each member state and potential changes from the pilot data collection to the actual implementation of the morbidity statistics in all member states.

7.3. THE NEED FOR A NATIONAL HEALTH INFORMATION SYSTEM

As mentioned before, there is currently no routine data linking between the various health information data sources in Belgium. Consequently, the results for each indicator rely on the data as delivered by the most relevant source. Due to the shortcomings of each data source, it can be argued that by relying on one source to detect the prevalence or incidence of a given condition, too many cases are potentially missed. In addition, data linkage is mainly project-based, implying that researchers spend a lot of time on inventorying what data is available, requesting the desired data, and linking the data. Lots of time is spent on reaching consensus with all the data holders, adapting the datasets for researchers and building the pipelines for the data exchange. Therefore, while discussing the project with the various stakeholders, the development of a National Health Information System was proposed repeatedly.

Such a national HIS could be the basis to create a combined or distributed database, which would be invaluable to multiple epidemiological projects. Linking or integrating these existing digital collections of various sources requires an identification of the data collections within Belgian health information system, mobilisation of the data holders and exploration the opportunities for a sustainable and longitudinal connection between these data sources. Previous reports (e.g. EU Expert Group on Health Systems Performance Assessment (HSPA), 2019) and projects (e.g. BELCOHORT project, aiming to explore the feasibility of setting up a population-based cohort in Belgium) demonstrated repeatedly the need for a more integrated system.

However, even though most of the stakeholders have expressed a certain enthusiasm regarding such Health Information System and agree that such a Health Information System would enable a deeper understanding of the public health status, none of the stakeholders seemed to be willing to start with the development of such a system. Key pitfalls seem to be issues regarding the financing of such a system along with the privacy regulations and imposed risks of combining multiple datasets. Discussions at Sciensano have started to further investigate the possibilities of such a Health Information System and the possible solutions to the aforementioned issues. As A recent project proposal was drawn up (AHEAD project) in order to address the lack of an integrated health information system in Belgium, by exploring the possibility of harmonizing health data that exist within the Belgian health information system and making these accessible for scientific exploitation and scientific valorization.

7.4. THE NEED FOR NATIONALLY REPRESENTATIVE PRIMARY CARE DATA

Difficulties have been encountered regarding the lack of availability of nationally representative primary care (diagnosis-based) data, as there is no national system in place that allows to (structurally) collect such data. Nevertheless, to arrive at representative diagnosis-based morbidity statistics in primary care, a more unified system that allows the collection, harmonization and use of these types of data is of great importance.

During the current COVID-19 pandemic, the national monitoring project *COVID-19 Barometer* was launched, in order to provide syndromic surveillance in GP practices and triage posts and, if necessary, provide feedback. Participation is open to all GPs. Automated queries in GPs' medical software packages can generate aggregated numbers on a defined set of diagnoses, encoded in the EMR. At the end of the day, an audit system automatically counts the number of diagnoses encoded and an e-form can be completed by the GP with the results from this audit. Participating GP's are asked to transfer data of their practice at least twice a week. This instrument enables rapid monitoring (real-time data) on a local scale. In addition, the tool enables the GPs to visualise their own data, and to compare themselves to other practitioners in their circle or region (compiled and anonymous data). These data can also be used for research, risk prediction and underpin (local) health policy decisions.

Currently, the COVID-19 Barometer collects data on a limited set of indicators; presumption or confirmation of COVID-19, viral syndrome, influenza and acute respiratory tract infections. In order to arrive at nationally representative diagnosis-based morbidity statistics in primary care, the set of indicators could be extended and participation to the Barometer should be encouraged. This requires investments in engaging the primary care givers and demonstrate the added value for their general practice. On the other end, resources are needed to maintain and improve the technical aspects of the Barometer, the database maintenance, data analysis and knowledge translation. By doing so, this relatively simple system can evolve into a system that can not only make nationally representative compiled primary care (diagnosis-based) data available, but also provide GPs with a feedback system, with tools for visualising data and internal evaluations. Sciensano could lead the coordination of this project, supporting its long-term strategic priorities in terms of its public service missions.

The Intego Network of General Practitioners, as described in Section 4.2.1, incorporates an automated individual health data collection, based on the Electronic Medical Record (EMR) of the patient as registered in their GP's practice. The collection of such individual EMR data allows the network to convey the incidence and prevalence of a multitude of diseases, as well as data concerning diagnostic tests or applied therapies in the registered GP practices. As only GP practices in the region of Flanders are included in this network, implementing this system nationwide would be the ultimate objective, allowing for the collection of nationally representative morbidity data in primary care data, still enabling the GPs to compare and evaluate their own data. As this concerns individual data, linkages with other health information sources is possible (see also Section 7.3 – The need for a National Health Information System) such as the Hospital Discharge Data or other administrative data sources, facilitating public health research and supporting health policy decisions. Different elements are already in place, the solid Intego network with the automated data collection and the technical pipelines are built in collaboration with healthdata.be and eHealth. In order to upscale this network, investments in engaging GP practices are needed, demonstrating the added value for them, and upgrading the different software applications so that they are compatible with the process of data retrieval by Intego. Despite this being a challenging scenario that requires long-term investments, a unified system can facilitate the valorisation of primary care data collections for evaluation, surveillance and research.

Acknowledgements

The project has received financial support from the European Commission, as identified by grant agreement number: 847046 – 2018-BE-MORBIDITY.

The authors would like to thank the following people, who contributed to the project. In alphabetic order: Ine Alaerts (Federal Public Service of Health, Food Chain Safety and Environment), Nathalie Bossuyt (Sciensano), Karin De Ridder (Sciensano), Robrecht De Schreye (Sciensano), Kris Doggen (Sciensano), Laurence Geebelen (Sciensano), Birgit Gielen (IMA/AIM), Pascal Meeus (RIZIV/INAMI), Yves Parmentier (RIZIV/INAMI), Françoise Renard (Sciensano), Rodrigo Ruz Torres (RIZIV/INAMI), Belinda ten Geuzendam (IMA/AIM), Bert Vaes (Intego), Eline Vandael (Sciensano), Melissa Van Bossuyt (Sciensano), Bert Van Den Bogaert (Sciensano), Johan Van der Heyden (Sciensano), Johan Vanoverloop (IMA/AIM), Peter Willemé (Federal Planning Bureau)

References

Drieskens S., Braekman E., Charafeddine R., Demarest S., Berete F., Gisle L. & Van der Heyden J. Health Interview Survey, Belgium, 1997 - 2001 - 2004 - 2008 - 2013 - 2018: Health Interview Survey Interactive Analysis. Brussels: Sciensano. <https://hisia.wiv-isp.be/>

EU Expert Group on Health Systems Performance Assessment (HSPA), 2019. Tools and methodologies to assess the efficiency of health care services in Europe.

Eurostat. (2014). Morbidity statistics in the EU. <https://doi.org/10.2785/52346>

Eurostat. (2018). Morbidity Statistics - Methodology. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Morbidity_statistics_-_methodology

KU Leuven. (2019). Intego. Retrieved from <https://intego.be/nl/Welkom>

Renard, F., Van Der Heyden, J., & Tafforeau, J. (2016). Morbistat_BE_FinalReport. Brussels.

Annexes

A. DATA SOURCE SELECTION

Indicator	Disease	Measure	Remark	Reference Period	Label	ICD-10 Chapter	ICD-10 Code	ICD-9 Code	ICPC-2 Code
P1	Diabetes mellitus	Incidence by Person		1 year	ENMDs	IV	E10-E14	250	T89-T90
P2	Diabetes mellitus	Period Prevalence		3 years	ENMDs	IV	E10-E14	250	T89-T90
P3	Dementia	Period Prevalence	incl. Alzheimer's disease	3 years	MBDo	V	F00-F03, G30	290, 290.10, 290.11, 290.13, 290.21, 290.40, 290.41, 290.8, 290.9, 294.10, 294.11, 294.20, 294.21, 331.0, 331.19, 331.82	P70
P4	Mental & behavioural disorders due to use of alcohol	Period Prevalence	incl. alcohol dependence	3 years	MBDo	V	F10	290	P15-P16
P5	Schizophrenia, schizotypal & delusional disorders	Period Prevalence		3 years	MBDo	V	F20-F29	295	P72
P6	Mood (affective) disorders	Period Prevalence		3 years	MBDo	V	F30-F39	301.12, 311	P73, P76
P7	Anxiety disorders	Period Prevalence		3 years	MBDo	V	F40-F41	-	P74, P79
P8	Parkinson's disease	Period Prevalence		3 years	DsNS	VI	G20	966.4, E936.4	N87
P9	Multiple sclerosis	Period Prevalence		3 years	DsNS	VI	G35	340	N86

P10	Epilepsy	Period Prevalence		3 years	DsNS	VI	G40-G41	345	N88
P11	Hypertensive diseases	Incidence by Person		1 year	DsCS	IX	I10-I13, I15	401-405	K86, K87
P12	Hypertensive diseases	Period Prevalence		3 years	DsCS	IX	I10-I13, I15	401-405	K86, K87
P13	Ischaemic heart diseases	Period Prevalence		3 years	DsCS	IX	I20-I25	410-414	K74-K76
P14	Acute myocardial infarction	Incidence by Episode		1 year	DsCS	IX	I21-I22	410	K75
P15	Acute myocardial infarction	Incidence by Person		1 year	DsCS	IX	I21-I22	410	K75
P16	Heart failure	Period Prevalence		3 years	DsCS	IX	I50	428	K77
P17	Stroke	Incidence by Person		1 year	DsCS	IX	I60-I64	430, 431, 432.0, 432.1, 432.9, 433.01, 433.11, 433.21, 433.31, 433.81, 433.91, 434.01, 434.11, 434.91	K90
P18	Cerebrovascular diseases	Period Prevalence		3 years	DsCS	IX	I60-I69	430-438	K90, K91
P19	Pneumonia	Incidence by Episode		1 year	DsRS	X	J12-J18	480.0, 480.1, 480.2, 480.3, 480.8, 480.9, 481, 482.0, 482.1, 482.2, 482.31, 482.32, 482.39, 482.40, 482.41, 482.42, 482.49, 482.81, 482.82, 482.83, 482.89, 482.9, 483.0, 483.1, 483.8,	R81 (Not perfectly matched to ICD-10)

								484.7, 485, 486, 487.0, 514, 517.0	
P20	Asthma	Incidence by Person		1 year	DsRS	X	J45-J46	493	R96
P21	Asthma	Period Prevalence		3 years	DsRS	X	J45-J46	493	R96
P22	Chronic lower respiratory diseases other than asthma	Period Prevalence	incl. COPD	3 years	DsRS	X	J40-J44, J47	490-492, 494, 496	R78, R79, R95, R99
P23	Chronic obstructive pulmonary disease (COPD)	Period Prevalence		3 years	DsRS	X	J44	490-492, 494, 496	R78, R79, R95, R99
P24	Alcohol liver disease	Period Prevalence		3 years	DsDS	XI	K70	571.0, 571.1, 571.2, 571.3, 572.2	D97
P25	Diseases of liver other than alcoholic	Period Prevalence		3 years	DsDS	XI	K71-K77	570, 571.40, 571.41, 571.42, 571.49, 571.5, 571.6, 571.8, 571.9, 572.0, 572.1, 572.2, 572.3, 572.4, 572.8, 573.0, 573.3, 573.4, 573.5, 573.8, 573.9	D97
P26	Diseases of liver	Period Prevalence		3 years	DsDS	XI	K70-K77	570, 571.0, 571.1, 571.2, 571.3, 571.40, 571.41, 571.42, 571.49, 571.5, 571.6, 571.8, 571.9, 572.0, 572.1, 572.2, 572.3, 572.4, 572.8, 573.0, 573.3, 573.4, 573.5, 573.8, 573.9	D97

P27	Rheumatoid arthritis	Period Prevalence		3 years	DsMSCT	XIII	M05-M06	714	L88
P28	Arthrosis	Period Prevalence		3 years	DsMSCT	XIII	M15-M19	715	L89-L91
P29	Osteoporosis	Period Prevalence		3 years	DsMSCT	XIII	M80-M82	733.00, 733.01, 733.02, 733.03, 733.09, 733.10, 733.11, 733.12, 733.13, 733.14, 733.15, 733.16, 733.19, 733.81, 733.82, 905.1, 905.2, 905.3, 905.4, 905.5, V54.21, V54.22, V54.23, V54.26, V54.27, V54.29	L95
P30	Renal failure	Period Prevalence		3 years	DsGS	XIV	N17-N19	583-586	U99
P31	Urolithiasis	Incidence by Person		1 year	DsGS	XIV	N20-N23	592, 594	U14, U95
P32	Intracranial injury	Incidence by Episode		1 year	IPEC	XIX	S06	850-854	N79, N80
P33	Intracranial injury	Incidence by Person		1 year	IPEC	XIX	S06	850-854	N79, N80

P34	Fracture of femur	Incidence by Episode		1 year	IPEC	XIX	S72	820, 821	L75
P35	Fracture of femur	Incidence by Person		1 year	IPEC	XIX	S72	820, 821	L75
PB36	Land transport accidents	Incidence by Episode		1 year	ECMM	XX	V01-V89	E800-E829	-
PB37	Land transport accidents	Incidence by Person		1 year	ECMM	XX	V01-V89	E800-E829	-
PB38	Accidental falls	Incidence by Episode		1 year	ECMM	XX	W00-W19	E880-E888	-
PB39	Accidental falls	Incidence by Person		1 year	ECMM	XX	W00-W19	E880-E888	-
PB40	Intentional self-harm	Incidence by Episode	incl. suicidal attempt	1 year	ECMM	XX	X60-X84	E950-E959	-
PB41	Intentional self-harm	Incidence by Person	incl. suicidal attempt	1 year	ECMM	XX	X60-X84	E950-E959	-
PB42	Complications of medical and surgical care	Incidence by Episode		1 year	ECMM	XX	Y40-Y66, Y69-Y84	E870-E876, E878-E879, E930-E949	-

PB43	Complications of medical and surgical care	Incidence by Person		1 year	ECMM	XX	Y40-Y66, Y69-Y84	E870-E876, E878-E879, E930-E949	-
-------------	---	----------------------------	--	--------	------	----	------------------	---------------------------------	---

Indicator	Disease	Pref. Data Source 1	Pref. Data Source 2	Pref. Data Source 3	Notes Data Source	DiagnosisType	ATC-code(s) EPS	ATC-code(s) Guidelines	Data Quality
P1	Diabetes mellitus	Permanent Sample of Health Insureds	Intego		Correction by calculated ratio from EPS data	Pseudo-diagnostic	A10A, A10B	A10	2. Good coverage & completeness
P2	Diabetes mellitus	Permanent Sample of Health Insureds				Pseudo-diagnostic	A10A, A10B	A10	2. Good coverage & completeness
P3	Dementia	Permanent Sample of Health Insureds	Intego			Diagnostic		N06DA, N06DX01	2. Good coverage & completeness
P4	Mental & behavioural disorders due to use of alcohol	Intego			Correction by calculated ratio from HIS data	Diagnostic			3. Doubtful completeness / methods
P5	Schizophrenia, schizotypal & delusional disorders	Permanent Sample of Health Insureds				Pseudo-diagnostic	N05AA, N05AB, N05AC, N05AD, N05AE, N05AF, N05AG, N05AH, N05AN, N05AX, N07XX06	N05A (except N05AN)	3. Doubtful completeness / methods
P6	Mood (affective) disorders	Intego				Diagnostic		N06A, N06C, N05AN except N06AX01 and N06AX02	3. Doubtful completeness / methods

P7	Anxiety disorders	Intego	National Hospital Stay Database		Preference to specific survey, validation/correction using HIS. MZG can deliver psychologic data?	Non-diagnostic			3. Doubtful completeness / methods
P8	Parkinson's disease	Intego			Correction by calculated ratio from EPS or HIS data	Diagnostic	N04AB, N04AC, N04B	N04	3. Doubtful completeness / methods
P9	Multiple sclerosis	Permanent Sample of Health Insureds				Pseudo-diagnostic	L03AB07, L03AB08, L03AX13, L05AA23, L05AA27, L05AA31, L05AA34, N07XX09		3. Doubtful completeness / methods
P10	Epilepsy	Intego			Correction by calculated ratio from EPS data	Diagnostic	N03	N03, N05BA, N05CD	3. Doubtful completeness / methods
P11	Hypertensive diseases	Intego				Diagnostic		C03AA C03AB C03AH C03AX01 C02CA04 C03BA C03DB C03EA C09BA02 C09BA03 C09BA04 C09BA05 C09BA06 C09BA07 C09BA08 C09BA09 C09BB C09DB C09DA02 C09DA03 C09DA04 C09DA06 C09DA07 C09DA01 C02AB01 C02AB02 C02AC01 C02AC02 C02AC04 C02AC05 C02DB02 C02DB03 C02DB04 C02DC01 C02DD01 C02DG01 C02KA01 C02KB01 C02KC01	2. Good coverage & completeness

								C02KD01 C02KX01 C09XA	
P12	Hypertensive diseases	Intego			Combined with EPS data during reference year to exclude lifetime prevalence	Diagnostic		C03AA C03AB C03AH C03AX01 C02CA04 C03BA C03DB C03EA C09BA02 C09BA03 C09BA04 C09BA05 C09BA06 C09BA07 C09BA08 C09BA09 C09BB C09DB C09DA02 C09DA03 C09DA04 C09DA06 C09DA07 C09DA01 C02AB01 C02AB02 C02AC01 C02AC02 C02AC04 C02AC05 C02DB02 C02DB03 C02DB04 C02DC01 C02DD01 C02DG01 C02KA01 C02KB01 C02KC01 C02KD01 C02KX01 C09XA	2. Good coverage & completeness
P13	Ischaemic heart diseases	Intego			Correction by calculated ratio from HIS data	Diagnostic		C01DA	2. Good coverage & completeness
P14	Acute myocardial infarction	National Hospital Stay Database	Cause of Death Registry		CDR data inclusion for calculation of total unique cases	Diagnostic			1. Good quality
P15	Acute myocardial infarction	National Hospital Stay Database	Cause of Death Registry		CDR data inclusion for calculation of total unique cases	Diagnostic			1. Good quality
P16	Heart failure	Intego			Correction by calculated ratio from EPS data	Diagnostic			3. Doubtful completeness / methods

P17	Stroke	National Hospital Stay Database	Cause of Death Registry		CDR data inclusion for calculation of total unique cases	Diagnostic			1. Good quality
P18	Cerebrovascular diseases	National Hospital Stay Database	Cause of Death Registry		CDR data inclusion for calculation of total unique cases	Diagnostic			1. Good quality
P19	Pneumonia	Intego	National Hospital Stay Database		Correction by calculated ratio from HDD data	Diagnostic			4. Doubtful quality
P20	Asthma	Intego			Correction by calculated ratio from HIS data	Diagnostic	R03DC01, R03DC03, R03DX05 - R03A, R03BA - R03	R03 (below 50 years of age)	2. Good coverage & completeness
P21	Asthma	Intego			Correction by calculated ratio from HIS data	Diagnostic	R03DC01, R03DC03, R03DX05 - R03A, R03BA - R03	R03 (below 50 years of age)	2. Good coverage & completeness
P22	Chronic lower respiratory diseases other than asthma	Intego			Correction by calculated ratio from HIS data	Diagnostic		R03	3. Doubtful completeness / methods
P23	Chronic obstructive pulmonary disease (COPD)	Intego			Correction by calculated ratio from HIS data	Diagnostic	R03BB, R03DA04 - R031, R03BA - R03	R03	3. Doubtful completeness / methods
P24	Alcohol liver disease	Intego			Using 'experiencing problematic drinking'- item HIS data to estimate subdivision proportion	Diagnostic			4. Doubtful quality
P25	Diseases of liver other than alcoholic	Intego			Using Pooled liver diseases HIS data to estimate subdivision proportion	Diagnostic			4. Doubtful quality
P26	Diseases of liver	Intego			Correction by calculated ratio from HIS data	Diagnostic			3. Doubtful completeness / methods

P27	Rheumatoid arthritis	Intego			Correction by calculated ratio from HIS data	Diagnostic	L04AA11, L04AA12, L04AB01, L04AB02 - L04AA12, A07EC01, A07EC02 - L04AA24, L04AB04, L04AB05, L04AB06, L04AB07		3. Doubtful completeness / methods
P28	Arthrosis	Intego			Correction by calculated ratio from HIS data	Diagnostic			3. Doubtful completeness / methods
P29	Osteoporosis	Intego			Correction by calculated ratio from HIS data	Diagnostic		A12A, G03XC, H05A, M05B	3. Doubtful completeness / methods
P30	Renal failure	Intego	European Dialysis and Transplant Association		Intego can deliver results on whole-stage prevalence. EDTA can only register end-stage prevalence (narrower definition, to be used as alternative/validation)	Diagnostic			2. Good coverage & completeness
P31	Urolithiasis	National Hospital Stay Database				Diagnostic			3. Doubtful completeness / methods
P32	Intracranial injury	National Hospital Stay Database	Cause of Death Registry			Diagnostic			2. Good coverage & completeness
P33	Intracranial injury	National Hospital Stay Database	Cause of Death Registry			Diagnostic			2. Good coverage & completeness
P34	Fracture of femur	National Hospital Stay Database				Diagnostic			1. Good quality

P35	Fracture of femur	National Hospital Stay Database				Diagnostic			1. Good quality
PB36	Land transport accidents	National Hospital Stay Database	Cause of Death Registry		Road Traffic Accident registry can be used for verification or correction. RTA not based on diagnostic data. NHSD expected to underestimate incidence	Non-diagnostic			4. Doubtful quality
PB37	Land transport accidents	National Hospital Stay Database	Cause of Death Registry		RTA registry does not allow to link multiple accidents to one person, IP not computable. NHSD delivers underestimation				4. Doubtful quality
PB38	Accidental falls	National Hospital Stay Database	Intego		NHSD only most severe cases, serious underestimation. Possible correction using HIS. Intego possible less severe cases - eg. Sprained ankle	Diagnostic			4. Doubtful quality
PB39	Accidental falls	National Hospital Stay Database	Intego		NHSD only most severe cases, serious underestimation. Possible correction using HIS. Intego possible less severe cases - eg. Sprained ankle	Diagnostic			4. Doubtful quality
PB40	Intentional self-harm	Sentinel Network of General Practices			Def. limited to suicidal attempt	Diagnostic			4. Doubtful quality

PB41	Intentional self-harm	Sentinel Network of General Practices			Def. limited to suicidal attempt	Diagnostic			4. Doubtful quality
PB42	Complications of medical and surgical care	National Hospital Stay Database	Cause of Death Registry	Intego	Most severe cases + deceased patients only. Narrow definition. Intego possible less severe - eg. Mild allergy	Diagnostic			4. Doubtful quality
PB43	Complications of medical and surgical care	National Hospital Stay Database	Cause of Death Registry	Intego	Most severe cases + deceased patients only. Narrow definition. Intego possible less severe - eg. Mild allergy	Diagnostic			4. Doubtful quality

B. MEETINGS WITH EXTERNAL STAKEHOLDERS

Date	Subject
10/04/2019	Task Force on Morbidity Statistics
25/04/2019	Internal kick-off meeting of the Morbidity Statistics pilot data collection
26/04/2019	Stakeholder meeting & project introduction with Intego Network of General Practitioners and their role in the Morbidity Statistics project
07/05/2019	In-house meeting on the previous phases of the Morbidity Statistics project
12/06/2019	Stakeholder meeting & project introduction with the FPS Health, Food Chain Safety & Environment regarding the role of the MHD and the NHSD
02/07/2019	Stakeholder meeting & project introduction with the IMA/AIM and the role of the EPS data in the Morbidity Statistics project
11/09/2019	In-house meeting regarding the data regulations and data access
26/11/2019	Stakeholder meeting with the Belgian Focal Point & introduction/clarification of the Morbidity Statistics project
12/12/2019	In-house meeting discussing the approach to achieving a more integrated Health Information System
11/06/2020	Stakeholder meeting (1/2) with FPS Health, Food Chain Safety & Environment & RIZIV/INAMI on Health Information System; examining the possibility of establishing a more structural link between health data sources
18/06/2020	Stakeholder meeting (2/2) with Statbel, IMA/AIM, Intego & Kruispuntbank van de Sociale Zekerheid on Health Information System; examining the possibility of establishing a more structural link between health data sources
02/12/2020	Meeting with Intego & Sciensano on the recommendations formulated in the final Morbistat Report

CONTACT

Brecht Devleesschauwer • T+32 2 642 50 35 • brecht.devleesschauwer@sciensano.be

MORE INFORMATION

—
Visit our website
>www.sciensano.be or
contact us at
>info@sciensano.be

Sciensano • Rue Juliette Wytsmanstraat 14 • Brussels • Belgium • T + 32 2 642 51 11 • T press + 32 2 642 54 20 •
info@sciensano.be • www.sciensano.be

Responsible editor: Christian Léonard, Managing director • Rue Juliette Wytsmanstraat 14 • Brussels • Belgium • D/2020/14.440/86