sciensano

BELGIAN NATIONAL BURDEN OF DISEASE STUDY

Methodological framework for cost estimation

January 2023, Brussels, Belgium

Sciensano

Epidemiology and public health • Health information

January 2023 • Brussels • Belgium

Deposit number: D/2023/14.440/4

–

Authors

Masja SCHMIDT

•

Vanessa GORASSO

•

Robby DE PAUW

•

Brecht DEVLEESSCHAUWER

Contact • burden@sciensano.be

Visit our website: https://burden.sciensano.be

Sponsors

The Belgian National Burden of Disease Study (BeBOD) is conducted as part of the Health Status Report project, and receives financial support of the National Institute for Health and Disability Insurance.

Acknowledgements

The Belgian National Burden of Disease Study (BeBOD) is an initiative of the Health information Service, Department of Epidemiology and Public Health, Sciensano. We would like to thank the following colleagues for their support and review: Finaba Berete, Leonor Guariguata, Sarah Nayani, Aline Scohy, and Johan Van der Heyden. We are also grateful to our partners, i.e., Statistics Belgium, Intego, the Belgian Cancer Registry, the European Register for Renal Replacement Therapy, the Federal Public Service Health, Food Chain Safety and Environment, and the Intermutualistic Agency.

Please cite this publication as follows:

Schmidt M., Gorasso V., De Pauw R., Devleesschauwer B., Belgian national burden of disease study. Methodological framework for cost estimation. Brussels, Belgium: Sciensano. Deposit number: D/2023/14.440/4

TABLE OF CONTENTS

1. General methodology 5

1.1. Cost of illness definition 5
1.2. Types of cost of illness studies 5
1.3. Productivity losses and patients' costs 6
1.4. Proposed approach and data sources 6
1.5. Collaboration with OECD 6

2. Inventory of linked databases 7

2.1. Linking health expenditures to diagnoses 7

3. General methodology for calculating incremental attributable costs 10

3.1. Background 10
3.2. Proposed general methodology 10

4. Specific methodologies by (linked) database 12

4.1. Intermutualistic Agency Echantillon permanent – Permanente steekproef 12
4.2. HISLink 12
4.3. Hospital discharge data (Technical cell – Cellule Technique) 12
4.4. Belgian Cancer Registry – Intermutualistic Agency 13
4.5. Intego 13

5. Health expenditure by disease based on the System of Health Accounts 15

6. References 16

Annex I – Methodology for calculating incremental attributable costs based on Gorasso et al., (2022) . 18
Annex II – Major Diagnostic Categories (MDCs) 20
Annex III – Variables to be included in Intego-IMA linkage 21

ABBREVIATIONS

AP-DRGs	All Patient Diagnosis Related Groups
APR-DRGs	All Patient Refined Diagnosis Related Groups
ATC	Anatomical Therapeutic Chemical Classification System
BCR	Belgian Cancer Registry
CBSS	Crossroads Bank for Social Security
DDD	Defined Daily Dose
EPR	Electronic Patient Record
EPS	Permanent sample
FPS	Federal Public Service (FPS) Health, Food Chain Safety and Environment
GLM	Generalized linear model
GP	General practitioner
GZSS	Gezondheidszorg – Soins de Santé
HIS	Belgian Health Interview Survey
HISLink	Linkage of Health Interview Survey Data with Health Insurance Data
ICD-10	International Classification of Diseases
ICPC	International Classification of Primary Care
IMA	InterMutualistic Agency
MCD	Minimum Clinical Data
MDC	Major Diagnostic Category
MHD-MZG-RHM	Minimum Hospital Data
MND	Minimum Nursing Data
MUG	Medical Urgencies Data
NISS	Social Security ID
OECD	Organisation for Economic Co-operation and Development
RIVM	Rijksinstituut voor Volksgezondheid en Milieu
RIZIV-INAMI-NIHDI	National Institute for Health and Disability Insurance
SHA	System of Health Accounts
TCT	Technische Cel – Cellule Technique

1. General methodology

1.1. COST OF ILLNESS DEFINITION

Cost of illness is a summary of the costs of a particular disease to society (1). When combined with data on morbidity and mortality, cost of illness studies can help to rank diseases according to their overall burden. In addition, cost of illness studies can help to identify the main cost components and their contribution to the total costs, so that decision-makers can develop targeted cost-containment measures (2). Cost of illness studies are also often used as an input for other health economic studies such as cost-effectiveness and cost-utility analyses. The latter are an important tool to support decisions on resource allocation in health care to ultimately improve the health of the population.

1.2. TYPES OF COST OF ILLNESS STUDIES

Cost of illness studies are often placed into different categories: they may be prevalence-based or incidence-based depending on the epidemiological data used, and may use top-down, bottom-up, or econometric approaches to estimate costs.

In **prevalence-based costs** of illness studies the direct costs and production losses attributed to all cases of a particular disease in a given year are estimated. The results of prevalence-based cost of illness studies are particularly useful for drawing attention to the burden of a particular disease and for planning cost containment policies (2). In **incidence-based** cost of illness studies on the other hand, direct costs and productivity costs associated with a given disease are calculated over a longer period of time, for example a lifetime, and assigned to the year that the disease first occurred. This approach is most useful for estimating potential savings associated with preventive measures, and for analysing the management of a disease from its onset until the end-of-life (2).

In addition, there are different ways to calculate costs. With the **top-down** approach, total national health care expenditures are allocated by type of care (e.g., hospital care, physicians' services) to different broad disease categories. These categories can for example be based on the International Classification of Diseases (ICD-10). All expenditures are allocated to the primary diagnosis, so that all costs add up to the total national health expenditure. Top-down cost of illness studies are typically prevalence-based (2). Examples of such top-down cost of illness studies starting from the total health expenditure can be found in the Netherlands (3) (see **Section 5**), Canada (4), and Germany (5).

In the **bottom-up** approach the average costs of treating a disease are calculated by multiplying the unit costs of all different treatment components by the average amount of utilization. The result is then multiplied by the prevalence of the disease to estimate total costs (6).

A third way to estimate costs 'attributable' to a disease is to compare costs of individuals with and without a disease by matching samples or using **regression analyses**. An example of a cost of illness study based on regression analyses is the one by Gorasso et al., (2022). The authors used a national health data set to compare a population with excess body weight (overweight and obesity) to a population with normal body weight, and controlled for age, sex, education, and behavioural risk factors including alcohol use, smoking, poor diets and physical activity. The resulting yearly incremental costs were multiplied by the prevalence of overweight and obesity to

achieve an estimate of the yearly healthcare costs and productivity loss associated with excess body weight in Belgium (7).

1.3. PRODUCTIVITY LOSSES AND PATIENTS' COSTS

Generally, the cost categories that are included in a cost of illness study are determined by the study perspective. Considering that the aim of the cost of illness study is to provide a summary of the costs of a particular disease to society, a **societal perspective** would often be most appropriate. This perspective implies that direct costs of treating the disease such as healthcare system costs for diagnosis, treatment and management of disease progression are included, but also patients' own costs (travel, over-the-counter medication), as well as indirect costs such as productivity loss resulting from time off employment (1,8).

As an example, a frequently cited problem with top-down cost of illness study is that they do not include costs categories that are not included in the national health care expenditures, for example costs associated with transport or informal care. Indeed, the size of these non-health costs may differ greatly by the different disease categories (2). Depending on the specific methodologies used and the availability of data, this problem may also occur in bottom-up and econometric cost of illness studies.

However, more narrow perspectives (e.g., health care system perspective) may be appropriate depending on the intended audience for the study and the way in which the study results will be used.

1.4. PROPOSED APPROACH AND DATA SOURCES

For the cost of illness study in the context of the Belgian National Burden of Disease Study, we propose a **prevalence-based, econometric approach to estimate the 'attributable costs' for different diseases in Belgium in a given year**. This implies that we will compare the health expenditure for people with and without a disease of interest. Considering that the main source of health expenditure data will be the InterMutualistic Agency (IMA) database, we will include **direct medical costs from the perspective of the healthcare payer** only. These include direct healthcare costs paid for in the context of the compulsory national scheme for health insurance and disability benefits, as well as official patients' copayments and supplements. In future iterations of the study, however, the perspective and scope might be expanded.

1.5. COLLABORATION WITH OECD

Recently, Sciensano started an informal collaboration with the OECD Expert Group on the Economics of Public Health (EGEPH). The EGEPH works on a project that aims to among other things improve healthcare expenditure estimates by training neural networks using individual-level data on use of healthcare services and related expenditure. The overall aim of the project is to improve the breadth and quality of the OECD modelling-based work on public health (see (24)).

In the context of the current collaboration, the OECD team will share the code to run the analyses with the Sciensano team. The Sciensano team will apply the code to the HISLink dataset to train the neural networks, and return the outputs from the analyses to the OECD team. By engaging in this effort, we will contribute to the work of the OECD, but we will also be able to use the tool to make our own health expenditure estimates to complement the BeBOD study.

2. Inventory of linked databases

2.1. LINKING HEALTH EXPENDITURES TO DIAGNOSES

To estimate the costs attributable to a certain disease using an econometric approach, we need data on health expenditures for individuals with and without the disease. Health expenditure data are available in the IMA database. However, this database does not contain explicit information on the health status of the individual. Therefore, the health expenditure data need to be linked to data from other databases that contain information on an individual's health status (i.e., diagnoses). To this end, an inventory of existing databases that link health expenditure and diagnosis was made. The results of this exercise are described in this section. Considering that the databases differ in structure and data availability, a description of the specific suggested methodology per database is also provided (**Section 4**).

2.1.1. Intermutualistic Agency (IMA) database

In Belgium, 99% of the population is covered by the compulsory health insurance. This insurance partly or in some cases fully covers the costs of a wide range of medical and paramedical services and pharmaceuticals. The 1% that is not covered are people whose administrative requirements, financial requirements, or both, have not been fulfilled. This calculation does not include irregular migrants and people covered by other insurance schemes (e.g., non-Belgians working for international organizations) (9).

The administrative management of the health insurance is carried out by seven health insurance organizations, the "mutualiteiten - mutualités". In 2002, an agency tasked with the collection and analysis of data from the seven health insurance organizations was founded: the InterMutualistic Agency (IMA). IMA manages a database that contains the Pharmanet and Gezondheidszorg – Soins de Santé (GZSS) datasets, as well as socio- demographic data for all Belgian citizens with (compulsory) health insurance (the 'Population' dataset). The IMA database is not a linked database in itself, but has been linked to other databases, such as those from the Belgian Health Interview Survey (HIS) and the Belgian Cancer Registry (BCR).

2.1.1.1. Pharmanet

Since 1996, the National Institute for Health and Disability Insurance (RIZIV-INAMI) collects data within the Pharmanet framework to monitor general practitioners' and specialist physicians' prescribing practices. Within the framework, data are collected by prescriber on the pharmaceutical supplies (e.g., magistral preparations, diabetic sterile syringes) delivered by public dispensaries (i.e., pharmacies) and stored in a database. The focus of the Pharmanet database is exclusively on reimbursed prescription drugs in ambulatory medicine delivered by public dispensaries. The database includes the costs (i.e., costs reimbursed by the health insurance, patient co-payments, and supplements) of these medications. The unique beneficiary identification number in the Pharmanet database is kept for a period of three years.

RIZIV-INAMI uses the Pharmanet data among other things to identify the occurrence of specific pathologies in the Belgian population. To this end, experts constructed a list of so called "pseudo-diagnoses", or "pseudo-pathologies" based on the delivery of drugs in the public pharmacies, using codes from the Anatomical Therapeutic Chemical Classification System (ATC codes). This system of alphanumeric codes was developed by the World Health Organization for the classification of drugs and other medical products. An individual in the database is marked as

being affected by a certain pseudo-pathology when the sum of the Defined Daily Dose (DDD) in the reference year is equal to or higher than ninety. The DDD is a unit of measurement defined as the assumed average maintenance dose per day for a drug used for its main indication in adults (10).

2.1.1.2. Gezondheidszorg – Soins de Santé

Since January 2014, the IMA database also contains a permanent healthcare dataset called GZSS. For all insured persons within the mandatory health insurance, this dataset contains details of their reimbursed healthcare provisions using nomenclature codes, which is a coded list of the healthcare provisions partially or totally reimbursed by the healthcare insurance. Information on reimbursed prescription drugs in hospitals pharmacies is also available.

2.1.1.3. Echantillon permanent – Permanente steekproef

For research purposes, the IMA created the permanent sample (EPS). The EPS is a sample of 1/40 of the population in the entire IMA database, with an oversampling of 1/20 of the population older than 65 years. A legal framework regulates the modalities for using the EPS to study and monitor health care consumption and expenditure in Belgium. Data are available from 2002 onwards. In contrast to the Pharmanet dataset, the EPS data represent a longitudinal dataset with a patient identifier that does not expire.

2.1.2. Linkage of Health Interview Survey Data with Health Insurance Data (HISLink)

In the HIS, Sciensano collects information on the health status, lifestyle and medical consumption of a representative sample of the general Belgian population, including community-dwelling older people. Information is also collected on a wide range of sociodemographic background characteristics. Interviews are carried out through a face-to-face interview and a self-complete questionnaire. The basic sample consists of 10,000 persons but oversampling of specific population groups is possible. By using weighting factors representative results can be calculated at the level of the total population. To date, the HIS has been organized in Belgium in 1997, 2001, 2004, 2008, 2013, and 2018. The next HIS will take place in 2023.

Recently, the data from the 2013 and 2018 HIS were linked with the health insurance data from the IMA database in the context of the Linkage of Health Interview Survey Data with Health Insurance Data (HISLink) project. The project was carried out by Sciensano in collaboration with IMA and Statistics Belgium, and funded by RIZIV-INAMI. Through individual linkage of the HIS data with the IMA database, some of the shortcomings of the health insurance data (no information on non-reimbursed health care use, no link between health care use and health needs, and scarce information on socio-demographic characteristics) could be overcome. This allowed to answer policy-relevant questions that cannot be answered by analysing each of the databases separately (11).

2.1.3. Hospital Discharge Data (Technical Cell – Cellule Technique)

The Belgian public health authorities collect records for all hospital stays in general hospitals in the Minimum Clinical Data (MCD). MCD registration for hospitalized patients was developed in the 1980s and became compulsory in 1990.

The MCD contain relevant clinical data (e.g., the primary and secondary diagnosis recorded for the hospital stay) and patients' demographic characteristics. The MCD are used to assign

hospitalized patients to Diagnosis Related Groups (DRGs). In 1995, All Patient DRGs (AP-DRGs) were used to group patients to allow comparisons between hospitals for financial purposes. In 2002, the AP-DRGs were replaced by the APR-DRGs (All Patient Refined DRGs, 3M HIS version 15.0) that take the severity of the illness into account.

In 2009, an integrated system for data collection called the Minimum Hospital Data (MHD-MZG-RHM) was launched. The Minimum Hospital Data Set integrates the MCD, the Minimum Nursing Data (MND) and the Medical Urgencies Data (MUG). These hospital discharge data are mainly collected as a tool for the measurement of hospital needs for public financing, and for the evaluation of the effectiveness and quality of hospital care. Other possible use cases for the data are internal management and to determine population needs through epidemiological studies.

The entity responsible for the management and processing of the hospital data is the Technische Cel – Cellule Technique (TCT) from RIZIV-INAMI and FPS Public Health (12). The TCT already linked and validated the Minimum Hospital Data from FPS Public Health with billing data obtained from the insurance institutions. The link between the two datasets was made via the identification number of the hospitalized patient with his insurance institution or via the social security ID (NISS) number delivered to the TCT, after irreversible coding, by the hospitals and the insurance institutions (13). The TCT publishes the average expenditures per hospital stay (minimum one overnight stay) for the different APR-DRGs. These expenditures include the costs reimbursed by the insurance institutions in the context of the mandatory health insurance scheme, and excludes patients' copayments (with a few exceptions), supplements, and costs reimbursed by additional insurances.

2.1.4. Belgian Cancer Registry – InterMutualistic Agency

Belgium has a nationally representative disease register for cancer, the BCR. The registry is a secondary data source for which data are collected from hospital records and other healthcare service records. The data on cancer diagnoses from the BCR has already been linked to the expenditure data in the IMA database in the context of an existing project within Sciensano. It should be noted that it concerns an ad hoc linkage of a specific sample of the total population of people living with cancer, with information on a limited number of variables.

2.1.5. Intego

The Intego network, operational since 1994, is an electronic patient record (EPR)-based network of 54 voluntarily participating general practitioner (GP) practices in Flanders which all use the same EPR software. The network is coordinated by the Academic Centre for General Practice at the KU Leuven and covers approximately 2% of the Flemish population. The Intego database contains information on diagnoses (primarily based on the International Classification of Primary Care (ICPC) coding system) and prescribed drugs. Aggregated results for the most common disorders can be explored online via https://intego.gbiomed.kuleuven.be/intego-apps/inc_prev_v0/ and https://www.intego.be/resultaten/tool.

Currently, Sciensano does not have access to linked Intego-IMA data. This link will therefore have to be established as part of this project (see **Section 4.5**).

3. General methodology for calculating incremental attributable costs

3.1. BACKGROUND

The basic methodology for calculating the cost of illness (i.e., calculating incremental attributable costs by comparing individuals with and without a disease) will be the same regardless of the database that serves as a source for health status information (i.e., diagnosis). An exception are the calculations based on the hospital discharge data (see **Section 2.1.3**), where disease-specific costs per hospital stay are available and calculating incremental attributable costs will not be necessary. The following section provides an outline of the basic methodology.

Previous cost of illness studies within Sciensano have made use of an econometric approach based on multistage regression to estimate the incremental or attributable costs of among others obesity and overweight (7) and musculoskeletal disorders (14). The approach uses multivariable regressions with negative binomial distribution and log link to compute the marginal attributable costs of a disease. Specifically, it concerns a generalized linear model (GLM). The GLM model with negative binomial distribution and log link is needed because costs (the dependent variable in the regression model) are not normally distributed. Indeed, some patients may have very high costs, and others may have zero costs, making the distribution right skewed and truncated, respectively (15). In addition, costs cannot be negative. The steps used in the previous studies (7) carried out within Sciensano to estimate the attributable costs for different diseases are presented in **Annex I**.

3.2. PROPOSED GENERAL METHODOLOGY

Based on the methodology outlined in **Annex I**, the following general, basic methodology is proposed for calculating incremental attributable costs for different diseases in the current project.

3.2.1. Descriptive analysis

As a first step, a descriptive analysis will be carried out to describe the study sample by group membership. That is, the age (groups), sex distribution, and preferential reimbursement status for the group of individuals with and without the disease will be described. Preferential reimbursement status is mainly determined by household income and is often used as a proxy for socioeconomic status. The results of this analysis can be presented in a table; an example of such a table can be found in Balu & Thomas (16). This step will give an overview of the sample characteristics that can be used to better interpret the results.

In a second step, the total healthcare expenditure by costs category, as well as the total costs (i.e., mean and 95% confidence intervals) for the total sample and the two groups will be calculated and presented in a table (see e.g., (7), Table 2). This step is needed to give an idea about how the costs are distributed over the different cost categories, and how this may differ from the group of individuals with the disease of interest compared to the group of individuals without the disease.

3.2.2. Incremental costs

The incremental costs will be calculated with a regression model and recycled predictions, as described in **Annex I**. The covariates included in the regression will be age, sex, region, and

socioeconomic status as evidenced by the preferential reimbursement status. These covariates are chosen because they are available in the IMA population dataset, and will therefore be available for all analyses, regardless of the database used in the linkage.

3.2.3. Uncertainty

To reflect prediction and survey uncertainty (only for diseases analysed with the HISLink), means and confidence intervals (CIs) will be computed via bootstrapping with 1000 replicates and 1000 Monte Carlo simulations drawn per replication (1000*1000 interactions) (see (7)).

3.2.4. Total costs estimate

As a last step, the incremental costs will be multiplied by the prevalence to obtain an estimate of yearly total costs.

3.2.5. Additional analyses

For diseases analysed with the HISLink, it is possible to also correct for behavioural characteristics and other sociodemographic characteristics in an additional analysis (see **Section 4.2.1**). In some cases, it may also be useful to correct for comorbidities or test the contribution of comorbidities to the incremental costs (see (7)). Whether these additional analyses are useful and feasible to conduct depends on the disease and linked dataset and will be decided on a case-by-case basis.

4. Specific methodologies by (linked) database

This section outlines the specific methodologies and the practical implementation of the methodology described in **Section 3** per (linked) database. This section also provides an indication of which database will be used for which type of diseases.

4.1. INTERMUTUALISTIC AGENCY ECHANTILLON PERMANENT – PERMANENTE STEEKPROEF

For conditions or diseases for which there are nomenclature codes or reimbursed drugs that are specific to this disease, with high prescription rates for the service or drug in question, the diagnosis data for the cost analysis will be derived from the IMA-EPS database. The Sciensano team has permanent access to the EPS data.

The incremental costs will be calculated following the general methodology outlined in **Section 3**.

4.2. HISLINK

For diseases for which there is a specific question included in the HIS, and a low risk that the question on this disease will lead to social desirability bias, the HISLink dataset will be used as the source of information on diagnosis.

4.2.1. Additional analysis

With the HISLink dataset, it is possible to include variables in the multivariable model that extend beyond those included in the IMA-dataset such as age, sex, reimbursement status, and region. Examples of these variables in the HISLink dataset include household education level, household income level, and behavioural characteristics such as smoking status, alcohol use, eating behaviour, and physical activity levels.

The additional variables from the HISLink dataset to be included in the final multivariable model will be selected with a "double-selection" approach. This means that the final model will include variables that predict either costs or disease status, identified with backward elimination (significant at 10% level).

4.3. HOSPITAL DISCHARGE DATA (TECHNICAL CELL – CELLULE TECHNIQUE)

For diseases that are in most cases treated in the hospital, the hospital discharge data (Minimum Hospital Data) will be used for diagnoses and costs.

The TCT publishes the average expenditures per hospital stay (minimum one overnight stay) for the different APR-DRGs. These expenditures include the costs reimbursed by the insurance institutions in the context of the mandatory health insurance scheme, and excludes patients' copayments (with a few exceptions), supplements, and costs reimbursed by additional insurances. The TCT also provides financial feedback by Major Diagnostic Category (MDC). Each MDC corresponds to an organ system or aetiology of a disease and is usually associated with a particular medical speciality. The DRGs are categorized into 25 MDCs, a pre-MDC group, and a rest group ('FF') (see **Annex II**).

In the context of the current project, we are interested in the average health expenditures per patient for different diseases based on the ICD-10, whereby patients are selected based on the primary diagnosis. The approach would therefore be to request these data by email from the TCT, where possible split up by region, sex, and age groups. The estimated health expenditure for the diseases for which we rely on the hospital discharge data would therefore not be based on incremental costs (i.e., comparing cases and controls), but rather represent the disease-specific costs incurred during hospitalization. These estimates will be combined with the prevalence of the disease in question to come to an estimate of the total direct medical costs associated with the disease.

Considering that the hospital discharge data concern average health expenditures per hospital stay, it is possible that expenditures not directly related to the disease of interest are also included, for example medication for diabetes. In addition, the data are limited to the costs incurred in hospital, meaning that possible ambulatory and other costs associated with the pathology are not included. It is worthwhile exploring a possible linkage of the hospital discharge data with IMA to also gain an insight in these costs, for example by adding up all ambulatory and other costs per person for the twelve months following the hospitalization. An additional further linkage with the Crossroads Bank for Social Security (CBSS) would allow to extend the cost calculations to costs associated with absenteeism (see e.g., (17) for an example of a cost of illness study using this triple-linkage). The resulting overview of the cost of illness could be presented on the disease level, as described above. Alternatively or additionally, the MDCs could be used, considering that this high-level categorization is of particular interest to policy.

4.4. BELGIAN CANCER REGISTRY – INTERMUTUALISTIC AGENCY

Since the Belgian Cancer Registry includes data on almost all patients diagnosed with cancer in Belgium, the costs of cancer will be assessed using the linked BCR-IMA database.

Considering the specific structure of the BCR database, a case-control study was proposed to evaluate the incremental attributable costs of different cancers. The average annual incremental cost of the general population will be compared to the costs incurred by the population that is affected by cancer in a specific cancer site. Every prevalent cancer patient with an active diagnosis in a predefined reference year will be included as well as a significantly acceptable proportion of the overall insured Belgian population as controls (1 case and 4 controls matched on age, gender, region, and preferential reimbursement status). A prevalent cancer patient is defined as a patient who is alive in the reference year and who had a cancer diagnosis in the ten years leading up to the reference year (i.e., 10-year prevalence).

The methodology for calculating incremental costs will follow the steps outlined in **Section 3.2**. This means that even though the cases are already matched on age, gender, region, and preferential reimbursement status, these variables will still be included in the general linear model, too. This will lead to more precise estimates of the incremental costs associated with different cancers.

4.5. INTEGO

For diseases that are known to be well-recognized in primary care, and for which people are frequently in contact with general practitioners, the Intego database is the most appropriate source of information for diagnosis.

The Intego dataset has been linked with the IMA database before for the years 2011-2015 in the context of PhD projects on prediction modelling at the KU Leuven. However, currently no linkage accessible to Sciensano exists.

Therefore, the Sciensano team will prepare a linkage scheme for a new Intego-IMA linkage. Variables that will have to be included from both databases (see **Annex III**) to perform the incremental costs analyses will be identified and an authorization request to the Information Security Committee will be prepared.

Considering that the Intego database has limited coverage, approximately 2% of the Flemish population, a matched case-control design based on linkage with IMA like the one used for the BCR is not possible. After all, one cannot be certain that the individuals not included in the Intego database, i.e., the controls that would be sourced from the IMA dataset, do not have the disease of interest. Therefore, data on both cases (individuals with disease of interest) and controls (individuals without the disease of interest) will have to be sourced from the Intego database. Costs will be calculated as the average costs in the reference year and four follow-up years.

Given the data source, the incremental attributable costs will be calculated for Flanders only, and not for the other regions. To calculate the total yearly costs for Belgium, the incremental attributable costs calculated for Flanders will be combined with the national prevalence data, based on the assumption that the incremental attributable costs will not differ substantially between the regions.

4.5.1. Example

According to the Intego database, 123,229 unique patients visited the general practices in the Intego network in 2015. This is the yearly contact group. Of these patients, 1,989 had a chronic alcohol abuse diagnosis (prevalence: 16.141 per 1,000 patients) (18). To calculate incremental costs of chronic alcohol abuse in the Intego population, the Intego data regarding health status (chronic alcohol abuse diagnosis or no chronic alcohol abuse diagnosis) of the individuals in the yearly contact group will be linked to their corresponding IMA data. In this case, this would lead to a maximum of 1,989 cases and 121,240 controls for the regression analysis.

5. Health expenditure by disease based on the System of Health Accounts

In the Netherlands, a top-down cost of illness study is conducted every four years by the Rijksinstituut voor Volksgezondheid en Milieu (RIVM) on behalf of the Ministry of Health, Welfare and Sport, based on health expenditure data supplied by Statistics Netherlands. The total health expenditures for a given year are broken down by age, sex, and diagnosis (dimensions of care demand) and sector, financing, care function (dimensions of care supply) (19). The diagnosis dimension consists of 94 specific groups and 25 rest groups (20). The results are published on a website (3).

In the Dutch study, total health expenditures are defined in four different ways, one of which follows the System of Health Accounts (SHA) definition. The SHA was developed by the Organisation for Economic Co-operation and Development (OECD) and is often used in international comparisons. It includes all costs associated with treatment and care, with the exception of nursing care (20). Providing estimates of national health care expenditure according to SHA is mandatory for all Member States of the European Union (21). The Belgian estimates are published on the Federal Public Service – Social Security website (21,22).

One of the aims of the current project was to explore the possibility of having a similar, top-down, cost of illness study for the Belgian context, based on the SHA. The outcome of this exploration was that it is probably not feasible to conduct similar research in Belgium. The sources that are used to construct the health accounts only contain limited data that can be used to link expenditure to diagnoses or pathologies. The data that is available mainly relates to screening and prevention, and not to curative care.

The administrative data collected in the Belgian health care system allow to know with quite a high level of precision how many procedures (e.g., MRIs) are reimbursed on a yearly basis and at what costs. However, with the exception of very specific procedures and nomenclature codes (e.g., in dental care), it is often hard or even impossible to determine the accompanying diagnosis.

The risk of not being able to allocate a large part of the health expenditure to diagnosis groups outweighs the benefit of having a top-down cost of illness study for the Belgian context. Therefore, it has been decided to focus on a regression-based analysis of incremental attributable costs for different diseases instead.

6. References

1. Cost of Illness [Internet]. YHEC - York Health Economics Consortium. [cited 2022 Sep 20]. Available from: http://yhec.co.uk/glossary/cost-of-illness/

2. Tarricone R. Cost-of-illness analysis. What room in health economics? Health Policy. 2006 Jun;77(1):51–63.

3. Kosten van ziekten | Volksgezondheid en Zorg [Internet]. [cited 2022 Aug 11]. Available from: https://www.vzinfo.nl/kosten-van-ziekten

4. Economic Burden of Illness in Canada, 2010 - Canada.ca [Internet]. [cited 2022 Aug 11]. Available from: https://www.canada.ca/en/public-health/services/publications/science-research-data/economic-burden-illness-canada-2010.html

5. Cost of illness [Internet]. Federal Statistical Office. [cited 2022 Sep 26]. Available from: https://www.destatis.de/EN/Themes/Society-Environment/Health/Cost-Illness/_node.html

6. Segel JE. Cost-of-Illness Studies—A Primer. :40.

7. Gorasso V, Moyersoen I, Van der Heyden J, De Ridder K, Vandevijvere S, Vansteelandt S, et al. Health care costs and lost productivity costs related to excess weight in Belgium. BMC Public Health. 2022 Sep 6;22(1):1693.

8. Part II: Economic Impact Analysis -- Cost of Illness: The Second of a Five-Part Series. :49.

9. Gerkens S, Merkur S. Belgium: Health System Review [Internet]. 2020. i–237 p. (Health Systems in Transition; vol. 5). Available from: https://eurohealthobservatory.who.int/publications/i/belgium-health-system-review-2020

10. WHOCC - Definition and general considerations [Internet]. [cited 2022 Sep 26]. Available from: https://www.whocc.no/ddd/definition_and_general_considera/

11. Linkage of Health Interview Survey Data with Health Insurance Data [Internet]. sciensano.be. [cited 2022 Sep 26]. Available from: https://www.sciensano.be/en/projects/linkage-health-interview-survey-data-health-insurance-data

12. Nationale Databank Medische Diagnose / Zorg & Kost [Internet]. [cited 2022 Sep 26]. Available from: https://tct.fgov.be/webetct/etct-web/

13. Algemene info & vragen voor de Feedback van de referentiebedragen | Nationale Databank Medische Diagnose / Zorg & Kost [Internet]. [cited 2022 Sep 26]. Available from: https://tct.fgov.be/webetct/etct-web/html/nl/fbho_faq.jsp

14. Gorasso V, Van der Heyden J, De Pauw R, Pelgrims I, De Clercq E, De Ridder K, et al. The Health and Economic Burden of Musculoskeletal Disorders in Belgium from 2013 to 2018 [Internet]. Rochester, NY; 2022 [cited 2022 Oct 18]. Available from: https://papers.ssrn.com/abstract=4166618

15. HERC: Analyzing Cost Data [Internet]. [cited 2022 Aug 11]. Available from: https://www.herc.research.va.gov/include/page.asp?id=analyzing-cost-data

16. Balu S, Thomas J. Incremental expenditure of treating hypertension in the United States. Am J Hypertens. 2006 Aug;19(8):810–6; discussion 817.

17. Traumatic brain injury hospitalizations in Belgium: A brief overview of incidence, population characteristics, and outcomes - PubMed [Internet]. [cited 2022 Oct 19]. Available from: https://pubmed.ncbi.nlm.nih.gov/36003627/

18. INTEGO [Internet]. [cited 2022 Sep 20]. Available from: https://intego.gbiomed.kuleuven.be/intego-apps/inc_prev_v0/

19. Kosten van Ziekten | RIVM [Internet]. [cited 2022 Sep 26]. Available from: https://www.rivm.nl/kosten-van-ziekten

20. Kosten van ziekten | Methode | Volksgezondheid en Zorg [Internet]. [cited 2022 Aug 11]. Available from: https://www.vzinfo.nl/kosten-van-ziekten/methode

21. Gezondheidsrekeningen (SHA) [Internet]. Federale Overheidsdienst - Sociale Zekerheid. 2018 [cited 2022 Aug 11]. Available from: https://socialsecurity.belgium.be/nl/cijfers-van-sociale-bescherming/statistieken-sociale-bescherming/gezondheidsrekeningen

22. Uitgaven voor gezondheidszorg [Internet]. Federale Overheidsdienst - Sociale Zekerheid. 2020 [cited 2022 Aug 11]. Available from: https://socialsecurity.belgium.be/nl/nieuws/uitgaven-voor-gezondheidszorg-14-05-2020

23. Yang W, Hamilton JL, Kopil C, Beck JC, Tanner CM, Albin RL, et al. Current and projected future economic burden of Parkinson's disease in the U.S. NPJ Parkinsons Dis. 2020;6:15.

24. Modelling work in public health: the OECD's SPHeP models - OECD [Internet]. [cited 2022 Oct 20]. Available from: https://www.oecd.org/health/modelling-work-in-public-health.htm

Annex I – Methodology for calculating incremental attributable costs based on Gorasso et al., (2022)

Unadjusted model

In a first step, the costs of patients with the disease and without the disease are compared by using an indicator variable for group membership as the explanatory variable (univariate regression, unadjusted model). The coefficient for the indicator variable represents the difference in cost associated with being a member of the experimental group (patients with the disease) compared to the control group (patients without the disease) (15). The unadjusted model also allows to examine the relative contribution of the four different cost categories (ambulatory, hospital, reimbursed medicines, and not specified) to the total (not: attributable) costs.

Selection of covariates

Since group membership is not random, in a second step additional dependent variables such as age, sex, and other factors (depending on the availability of data) will be added to the model. The independent variables or covariates included in the final multivariable model used to calculate the incremental costs are selected from a group of candidate variables by means of 'double selection'. The double selection approach ensures that the final model only includes variables that either predict the dependent variable (costs) or the independent variable (group membership) (significant at the 10% level).

Calculating incremental costs

In a third step, incremental costs will be calculated at the individual level (patient) using the method of recycled predictions, also called direct standardisation or g-computation. This method allows to estimate the marginal effect from the disease on the health expenditure. Specifically, the coefficients of the final regression model are used to make the following predictions:

- Predicted healthcare costs for each individual given their observed disease status (group membership)
- Predicted healthcare costs for each individual assuming everyone is healthy (i.e., does not have the disease of interest)

The predictions will then be used to calculate the incremental costs on the individual level (difference between the first and second prediction).

Transforming the individual incremental costs to a survey-weighted average

For analyses based on the HISLink, the attributable costs of the disease will be calculated as the survey-weighted average of the yearly individual incremental costs calculated in step three. This step also includes computation of means and confidence intervals by means of bootstrapping and Monte Carlo simulations, to reflect uncertainty.

Total cost of illness based on prevalence and incremental costs

Total costs will be calculated by multiplying the incremental costs calculated in step four with the estimated prevalence of the disease on the 1st of January of the year (e.g., 2018, the year in which the last HIS took place) for which the calculations were made.

Relative contribution of other chronic conditions

Depending on the availability of data on, the analysis can be extended by exploring the relative contribution of other chronic conditions to the incremental costs of the disease of interest (see, (7)). For practical reasons, these conditions are not included as candidate variables for the multivariable model.

Indirect costs

Depending on the data available, indirect costs based on reported absence from work (absenteeism) and the average daily wage in the Belgian working population will also be calculated. This is for example possible in analyses based on HISLink.

Annex II – Major Diagnostic Categories (MDCs)

Table A1 • List of Major Diagnostic Categories used by the TCT

Diagnostic Category	Description
0	Pre-MDC
1	Diseases and Disorders of the Nervous System
2	Diseases and Disorders of the Eye
3	Diseases and Disorders of the Ear, Nose, Mouth And Throat
4	Diseases and Disorders of the Respiratory System
5	Diseases and Disorders of the Circulatory System
6	Diseases and Disorders of the Digestive System
7	Diseases and Disorders of the Hepatobiliary System And Pancreas
8	Diseases and Disorders of the Musculoskeletal System And Connective Tissue
9	Diseases and Disorders of the Skin, Subcutaneous Tissue And Breast
10	Diseases and Disorders of the Endocrine, Nutritional And Metabolic System
11	Diseases and Disorders of the Kidney And Urinary Tract
12	Diseases and Disorders of the Male Reproductive System
13	Diseases and Disorders of the Female Reproductive System
14	Pregnancy, Childbirth And Puerperium
15	Newborn And Other Neonates (Perinatal Period)
16	Diseases and Disorders of the Blood and Blood Forming Organs and Immunological Disorders
17	Myeloproliferative DDs (Poorly Differentiated Neoplasms)
18	Infectious and Parasitic DDs (Systemic or unspecified sites)
19	Mental Diseases and Disorders
20	Alcohol/Drug Use or Induced Mental Disorders
21	Injuries, Poison And Toxic Effect of Drugs
22	Burns
23	Factors Influencing Health Status and Other Contacts with Health Services
24	Human Immunodeficiency Virus Infection
25	Multiple Significant Trauma
FF	Rest group

Annex III – Variables to be included in Intego-IMA linkage

Table A2 • Variables to be included in the Intego-IMA linkage

Variable	Code	Description	Source	Unit
Sociodemographic characteristics and health status				
Age	Based on PP0015 (year of birth)	Age as recorded in the reference year	IMA	Years
Sex	PP0020	Sex as recorded in the reference year	IMA	0:Male 1:Female
Death	Based on PP0040A	Death in reference year	IMA	0:No 1:Yes
Reimbursement status	MAJOR_COVERAGE_YN	Preferential reimbursement based on PP0030 (*code titulaire 1 – code gerechtigde)*	IMA	0:No 1:Yes
Socio-economic status	Based on PP1003 (social situation) and PP0030 (*code titulaire 1 – code gerechtigde)*	N/A	IMA	1: active manual labor (including the unemployed and people living with a disability) 2: active employee (including the unemployed and people living with a disability) 3 : statutory agents of the public service 4: active self-employed persons, including self-employed persons living with a disability, starters and

Variable	Code	Description	Source	Unit
				assimilated self-employed persons
Presence of disease of interest	N/A	GP visit for disease of interest in the reference year based on ICPC-Codes	Intego	String based on ICPC-Code
Health expenditure and healthcare utilization				
Aggregated health expenditure	Based on SS00060 (reimbursed costs), SS00160 (personal contributions) and SS00165 (supplements). Including 100% per diem (fixed fees in case of a hospitalization/nursing day price). [a]	Total individual health expenditure in reference year and four follow-up years	IMA	Euro
Health expenditure, reimbursed	SS00060. Including 100% per diem (fixed fees in case of a hospitalization/nursing day price)	Reimbursed individual health expenditure in reference year and four follow-up years	IMA	Euro
Health expenditure, personal contribution	SS00160. Including 100% per diem (fixed fees in case of a hospitalization/nursing day price)	Individual health expenditure in terms of personal contributions in reference year and four follow-up years	IMA	Euro
Health expenditure, supplements	SS00165. Including 100% per diem (fixed fees in case of a hospitalization/nursing day price)	Individual health expenditure in terms of supplements in reference year and four follow-up years	IMA	Euro

Variable	Code	Description	Source	Unit
Health expenditure category	PROCEDURE_AH_CAT	Variable that indicates health expenditure by function	IMA	1: Ambulatory care 2: Hospital care 3: Other 4: Pharmaceuticals [b]

[a] *The health expenditure variables do not include fixed fees related to hospitalizations. We will ask IMA to include these costs in these variables if possible.* [b] *The PROCEDURE_AH_CAT variable does not contain information on pharmaceuticals. However, we will ask IMA to provide this category based on the Farmanet database if possible.*
GP, general practitioner; IMA, InterMutualistic Agency