

STUDY PROTOCOL

Health Interview Survey 2023

Authors : HIS 2023 team

Contact person : Rana Charafeddine (rana.charafeddine@sciensano.be)

Contents

1. GENERAL INFORMATION	4
1.1. Coordinates of Sponsors	4
1.2. HIS team members	4
1.3. Subcontractors	4
2. OBJECTIVE OF THE STUDY	5
2.1. Description of overall objective	5
2.2. Description of specific objectives	5
3. SCIENTIFIC RELEVANCE OF THE STUDY	6
3.1. Scientific background	6
3.2. Conceptual framework	6
3.3. Public health relevance	9
4. METHODS	10
4.1. Study design	10
4.2. Sampling methodology	12
4.3. Instruments	14
4.4. Fieldwork procedures	19
4.5. Data flow and data management	24
4.6. Data analysis	28
4.7. Methodology used to improve the quality of the data (quality controls)	31
4.8. Software development	34
4.9. Website	35
4.10. Procedures to obtain the HIS micro data	36
5. COMMUNICATION OF THE RESULTS AND REPORTS	37
5.1. Reporting mechanism	37
5.2. Publication plan: peer-reviewed publications and others	37
5.3. Other forms of communication of results	37
6. RISK AND BENEFITS FOR PARTICIPANTS (ETHICS)	39
7. ORGANISATION OF THE RESEARCH PROJECT	42
7.1. Starting and completion date	42
7.2. Subcontracting Statbel	43
7.4. Scientific review	43
7.5. Propriety rights of study material and results	44
7.6. Client satisfaction	44
7.7. Archiving process	45
8. REFERENCES	46

Abbreviations:

AIM-IMA	Inter-Mutuality Agency (in FR-NL)
BHIS	Belgian Health Interview Survey
CAPI	Computer Assisted Personal Interview
CASI	Computer Assisted Self Interview
EHIS	European Health Interview Survey
EU	European Union
HES	Health Examination Survey
HISIA	Health Interview Survey Interactive Analysis
HIS 2023	Health Interview Survey organised in 2023
INAMI–RIZIV	National institution for social security (in FR-NL)
OECD	Organisation for Economic Co-operation and Development
NR	National Register
PAPI	Paper And Pencil Interview
RCM	Résumé Clinique Minimum
RIM	Résumé Infirmier Minimum
RPM	Résumé Psychiatrique Minimum
RP	Reference Person (of a household)
SAS	Statistical Analysis Software
UN	United Nations
WHO	World Health Organisation
XX	stands for the 2 letters of identification of the specific HIS modules

1. General information

1.1. Coordinates of Sponsors

The national Health Interview Survey (HIS) is commissioned and co-financed by the different Belgian authorities competent in the field of public health. The assignment to carry out the Health Interview Survey (HIS) in 2023 was determined in the framework of an Inter-ministerial Agreement between the Belgian Federal State and the Authorities (Regions and Communities) defined by the articles 128, 130 and 135 of the Constitution. The Inter-ministerial Agreement Protocol¹ was concluded on November 2021.

All cabinets and administrations of the Ministries inclined in the organisation of the HIS 2018 are represented in a formal “Commission of Commissioners”. The Commission meets periodically – at least twice a year or upon request – with the HIS project leader to monitor the progression of the survey.

The Commission of Commissioners is chaired by:

Mr Paul De Bock
Ministry of Public Health
Eurostation II – Place Victor Horta, 40/10 - 1060 Brussels

1.2. HIS team members

The team devoted to carrying out the HIS are as follows:

Project leaders:

Johan Van der Heyden and Stefaan Demarest

Scientific members (alphabetic):

Finaba Berete
Elise Braekman
Rana Charafeddine
Sabine Drieskens
Lydia Gisle
Lize Hermans
Christina Stabourlos
Pierre Smith

Administrative member:

Ledia Jani

1.3. Subcontractors

The HIS 2023 data collection phase is carried out by Statbel, the Belgian statistical office under conditions stipulated in an outsourcing contract².

¹ Stored on Sciensano server: [Protocol HIS2023 signed.pdf](#)

² Stored on Sciensano server: \\sciensano.be\fs\1140_DATA\HIS\HIS2023\Statbel

2. Objectives of the study

Health information and research was defined during the 43rd World Health Assembly (1) as “a process for obtaining systematic knowledge and technology that can be used for improvement of the health of the individuals or groups of population”. Health information can thus be considered as one of the tools to be used for health promotion and disease prevention. The HIS is thus a powerful framework for a rational policy decision-making process (2).

2.1. Description of overall objectives

The main objective of the HIS is to measure the health status of the population in Belgium, accounting also for the three regional sub-populations (in Flanders, Wallonia and Brussels-Capital). The HIS is designed to obtain information on people’s health experience, their attitudes and health-related behaviours, the extent to which they use health care facilities and their use of preventive health and social services.

Health surveys provide one possible channel through which health-related information can be obtained. The added value here is the horizontal approach of the data collection: several types of information (health status and determinants, personal characteristics, health consumption, etc.) are collected simultaneously from the same individuals. The outcome is a global picture of the population’s health that allows identifying priority domains for strategic interventions. In addition, because the data are gathered periodically over time, changes in health status as well as effects of health policies and interventions can be monitored. Last but not least, health surveys allow obtaining health information from a representative sample of the population, including people that cannot be reached through the health services.

The ultimate goal of the HIS is thus to inform health authorities and stakeholders on various aspects of health in the population, but also to influence policy and health programs with surveillance data and provide a rich database to the scientific community for in depth research activities. The information collected via the HIS is not only useful at regional, community and national level but also for international instances such as Eurostat, WHO, UN and OECD. Additionally, data from the HIS is used to provide the EU with the necessary information for the mandatory European Health Interview Survey (EHIS).

2.2. Description of specific objectives

More specifically, the health survey pursues the following objectives:

- The identification of health problems;
- The description of the health status and health needs of the population;
- The estimation of prevalence and distribution of health indicators;
- The analysis of social (in)equality in health and access to health services;
- The study of health consumption and its determinants;
- The study of possible trends in the health status of the population;
- And the contribution to the evaluation of specific public health programmes.

3. Scientific relevance of the study

3.1. Scientific background

National health surveys collect information on health issues by means of structured standardised interviews and/or examinations in a representative sample of the general population in order to support evidence-based policy decisions. To gather this type of data, most European countries conduct surveys that are exclusively oriented towards health-related items, as the case of the Belgian Health Interview Survey (BHIS). Other countries carry out multipurpose surveys that include a module on health among a myriad of other different topics. The advantage of an exclusive health-oriented survey over a multipurpose survey is that it allows a broader and in-depth investigation of the health issues at stake, as there are no other competing subjects under study (3). In Belgium, the traditional multipurpose surveys carry only a very limited number of questions pertaining to health, and miss some important aspects needed to guide the authorities in their decision making. The BHIS is thus considered as the main reference in terms of health statistics in Belgium.

The BHIS gathers health information by addressing the questions directly to the population, through a face-to-face interview and a self-administered paper questionnaire. In this sense, the data collected depends on the individuals' self-perception and is regarded as "subjective" (as opposed to medical records, for instance).

As was done for the HIS 2018, the HIS 2023 will also be linked with administrative data: individual data on the use of healthcare services and reimbursed medicines will be extracted from the registers of the Inter-mutuality Agency (AIM-IMA) and linked at an individual basis to the data obtained through the BHIS in order to obtain complete records. The aim of this linkage is twofold: first, the mutuality data ensures some "objectivity" regarding the factual information gathered, and second, it alleviates the burden of questioning survey participants about health topics that are available from other sources. The linkage of the administrative data to the HIS data will also be presented in a stand-alone study protocol.

A new feature of the HIS 2023 is the organisation of the Oral Health Data Registration and Evaluation Survey (OHDRES) with a subset of the HIS 2023 participants. The overall aim of OHDRES is to collect relevant and valid information on the oral health of the Belgian population. The survey consists of 3 parts: 1) an oral health module integrated in the HIS 2023 face-to-face questionnaire, 2) an oral health examination carried out by a qualified dentist, 3) a linkage of the data obtained from the two first parts with selected data from the national health insurance obtained through AIM-IMA. A separate protocol is available for this survey³ and it will submit its own ethical approval.

3.2. Conceptual framework

The BHIS covers five domains of investigation that constitute its conceptual framework:

- The health status and well-being of the population
- The lifestyle and behaviours that are determinants of health
- The health literacy and use of medical prevention

³ Stored at : \\sciensano.be\fs\1140_DATA\HIS\HIS2023\Oral health survey\Study protocol

- The use of health- and social services
- The physical and social environment related to health

Within each of these broad domains, specific topics are subject to investigation, in function of their relevance in terms of national and regional surveillance, and need for public health decision making.

3.2.1. Health status and well-being

Measuring the health status of a population is necessary according to the WHO definition of health and the global approach of health problems. An instrument such as the HIS is crucial to complement the information collected by health care providers (only available for those that attend health services), health- and care-related registries and vital statistics. The HIS theoretically allows assessing the global health status of the population, and can expand the enquiry beyond specific diseases and health problems. This difference is described in the literature as the distinction between 'health status' and 'state of the health' (4).

Despite the fact that the main subject of the HIS is "health" (as opposed to disease) and that the WHO recommends adopting a positive approach of the concept, most of the domains investigated in the BHIS address ill-health and disorders. A positive conceptual framework is indeed difficult to implement due to a lack of clear definitions of "good health" and suitable instruments to measure it (5;6). Nevertheless, some BHIS indicators can be regarded as having a bipolar value, as for instance: good/poor self-perceived health, psychological well-being/distress, good/poor quality of life, vital energy/exhaustion, high/low life satisfaction, etc. In terms of health measurement, one must note that the BHIS mainly focusses on the chronic conditions, not acute or infectious diseases. Due to their long duration, chronic disorders and conditions have a greater impact on health expenditures and represent a higher burden at the population level than acute ones. Next to the health disorders themselves, the BHIS also takes into consideration their actual consequences on the functional status of individuals, through an assessment of the disability level, for instance.

Finally, next to the *physical* dimension of health, disorders and disability, the BHIS also addresses some aspects of *mental* health and disorders. This is crucial, as the survey is the only way to gather global information on the psycho-emotional status at population level.

3.2.2. Lifestyle and health behaviours

Lifestyle is conceptualised as a determinant of health: some behaviours contribute to preserving a good state of health, preventing specific conditions and improving others. At the opposite, certain habits are harmful, especially if they appear in an excessive or chronic manner. Yet health behaviours related to health (ex. physical activity, nutritional habits, alcohol, tobacco and drug consumption...) are linked to the values and priorities of each person, to the opportunities and constraints inherent to their culture as well as to their socio-economic position. Health behaviours are shaped by social learning and interpersonal interactions. It is thus misleading to believe that a specific lifestyle is solely governed by the individual's willingness and decision to adopt certain behaviours (deterministic approach).

Nonetheless, lessons from the past have shown that changing people's lifestyle is the main source of improvement in population health, before and above the progress made in the field of medicine. This is why today health promotion programmes are still one of the most important components of public health policy. It is thus essential for decision makers and institutions in charge of implementing health promotion to regularly measure the prevalence of health behaviours at the population level, their

distribution in specific sub-groups and their trends over time. Such data are imperative for the evaluation of programmes and policies. Of course, health surveys do not provide evidence of a causal relationship between a specific programme and a change in behaviour over time, but they remain useful tools for monitoring behaviour trends in the population.

3.2.3. Medical prevention and health literacy

Clinicians have long understood that preventive medicine plays a major role in maintaining good health. The advantage of preventive medicine has become increasingly apparent in the past 50 years. Prevention deeply modified the way to solve medical problems such as infectious diseases (with the immunisation programmes, for example). More recently, early disease detection has also become an essential component of preventive medicine, with striking results as far as morbidity and mortality are concerned (7).

As a result, public health policy has progressively been enlarged from the mere management of health expenditures to the development of strategies that aim at improving the health of the population on the whole. This perspective involves adopting actions at the level of biological factors, environmental aspects, individual behaviours, but also at the level of the health services in both their curative and preventive components (8). From a conceptual point of view, three scopes can be distinguished in the field of preventive medicine (8):

- Primary prevention: actions taken to abrogate the cause of a disease in order to avoid emergence of new cases.
- Secondary prevention: early detection and treatment of a specific disease before the apparition of the clinical symptoms and the complications.
- Tertiary prevention: it is not strictly speaking prevention of diseases, but rather trying to limit their consequences.

The HIS examines some actions in the primary and secondary prevention scopes. Priority actions in the domain of preventive medicine are chosen for the BHIS on the basis of several criteria: the frequency of the disease, the importance of the problem at the individual and societal level and the efficacy of the preventive methods.

Beside the actions taken in the field of medical prevention, the population should have a certain ability to obtain, understand and use health information in order to make appropriate decisions in terms of health care, use of medical drugs, health screening and preventive measures or compliance to treatment. This is referred to as “health literacy” – which is having the capacities, skills, knowledge and motivation to access, understand, (...) and apply relevant health information in different forms, in order to make judgments, take decisions and act in everyday life in the domains of healthcare, disease prevention and health promotion to improve quality of life (and health) throughout the life course. Stated differently, limited health literacy implies a lack of knowledge or miscomprehension about the bodily function and the nature and causes of disease. As a consequence, the lifestyle to health outcome relationship is not correctly perceived. A facet of health literacy is approached for the first time in the BHIS 2018.

3.2.4. Use of health and social services

So-called “health consumption” data is an essential part of the health information system for adapting available resources to the needs of the population. Health consumption covers three main domains:

ambulatory care, institutional care and medicines. Two different methods are usually available to measure health consumption at population level: health service statistics and health surveys. Several data sources are accessible in Belgium regarding health service registers, for instance, the reimbursement of expenses for medical care (data from INAMI–RIZIV, insurance funds), the reimbursement of expenses for prescribed medical drugs (data from Pharmanet) and hospital discharge records (data from RCM, RPM, RIM).

It is generally considered that health service statistics are more reliable (objective) than information gathered through health surveys. Indeed, health information collected from survey participants is thought to be distorted through memory bias, social desirability bias, as well as lack of knowledge on certain health issues. Nonetheless, health surveys represent the only source of polyvalent information, where data are collected on different health aspects such as diseases, disabilities, lifestyle and socio-demographic characteristics. Nevertheless, to take advantage of the benefits of both data sources (allying diversity of information and objectivity of health consumption), the BHIS 2023 will use the registers' data regarding reimbursed medical care and medicines and link them on an individual basis to the data collected through the survey.

Health services and related social services for which no reimbursement is awaited, and that are thus not registered, are examined in the BHIS 2023. This pertains to non-conventional therapies, non-prescribed medical drugs, etc.

3.2.5. Physical and social environment

Health and ill-health are also regarded as being determined to some extent by the physical environment, as well as related to social insertion. It is now well documented that the health status is linked to the social and economic situation of individuals. This can be assessed through concepts such as the accessibility to health care, social support, environmental nuisances, and violence, and through the analysis of health inequalities based on socio-economic factors. Hence, the BHIS takes into account a number of environmental and social health determinants.

3.3. Public health relevance

Health related estimates are useful for informing country policy and planning in the domain of public health. Public health is the science and art of preventing disease, prolonging life and promoting health through the organised efforts and informed choices of society, organizations, public and private, communities and individuals (9). Today, the HIS constitutes an important source of information for people in charge of setting up public health policies. It provides a solid ground on which those policies can be constructed. The HIS can also prove to be an interesting tool in the evaluation process of the public health programmes.

4. Methods

4.1. Study design

4.1.1. Study type

The national Health Interview Survey – HIS – is a cross-sectional epidemiologic study carried out in the general population of Belgium. It has been repeated periodically over time since the first edition in 1997. The present edition, 7th of the kind, is launched in 2023.

4.1.2. Study period

The Inter-ministerial Protocol Agreement for financing the HIS 2023 covers 3 calendar years. Each year corresponds to a particular phase of the research:

2022	Preliminary phase: conceptual, questionnaire development, pre-test and adjustment, preparation of the fieldwork...).
2023	Fieldwork: data collection phase (01/01/2023 to 31/12/2023) and encoding.
2024	Data cleaning and management, data analysis and reporting, dissemination. This phase can be prolonged beyond the end of 2024.

4.1.3. Target population

Belgian residents

The main objective of the HIS – i.e., to give a description of the health status and related behaviors of the population in Belgium – leads to the broad definition of the target population as “all people residing in Belgium, regardless their place of birth, nationality or any other characteristic”.

4.1.4. Sampling frame

National Register

The sampling frame consists of all households listed in the National Register (NR).

Practically, a household includes all people that are declared living at the address of a reference person (attributed field code = 1). However, some people live in collective households (e.g. homes for the elderly, prisons, religious convents...) and there is no “reference person” (RP) in these cases. So in order to include institutions for the elderly in the sampling frame, individuals belonging to collective households (attributed field code = 20) receive the field code ‘1’ (that is, these codes are ‘artificially’ changed to RP = 1) and are considered as belonging to a one-person household.

Exclusion rules

People who are listed in the NR without any physical address (mostly people living abroad, hence even not included in the target population) are excluded *a priori* from the sampling frame. They are identified in the National Register through the field of the statistical sector (starting with “Z” or “9”).

It should be noted that for practical reasons, exclusion rules were applied to people living in:

- an institution (including psychiatric institutions), with the exception of people living in homes for the elderly, nursing homes and psychiatric nursing homes,
- a religious community or cloister that shelter more than 8 persons,
- a prison,

The exclusion process takes place *a posteriori*, when the interviewer identifies that the selected one-person household (in this case) belongs to one of the categories mentioned above.

Quarterly sampling

The sampling procedure in the NR is carried out every quarter. Hence, it is applied 4 times during the fieldwork phase, during the month that precedes the start of each quarter. Households that were selected in a former quarter are excluded from the sampling frame of the following quarters. Quarterly sampling in the NR is performed to ensure that the selected households correspond as much as possible to the real situation in the field.

Updates

Even though the NR is updated on a weekly basis and the sampling is due every 3 months, it can happen that the administrative composition of a household differs from its real, actual composition. As all the members of a selected household are part of the sampling frame, the rule in this case is to add or retrieve household members from the list established on the information from the NR, to conform to the real situation. The newly added people will not however be included in the sample as the constitution of the sample is based on the data in the NR.

The list should also include household members who are temporarily absent (e.g. students who spend most of their time in student accommodations) and older people (65 years and older) who are 'officially' registered at the household address, but live in a home for older people, a nursing home for older people or a psychiatric nursing home. However, if the reference person is no longer living at the selected address, the household is no longer part of the sample.

4.1.5. Study population

The study population – the population that is actually reached via the sampling frame – does not perfectly cover the target population (= all people of Belgium). The following categories of individuals are included in the target population, as defined above, but are not included in the study population:

1. People living in Belgium, but who are not listed in the NR: homeless, illegal immigrants, etc.
2. People, residing in:
 - an institution (including psychiatric institutions), with the exception of older people living in old people's homes, nursing homes and psychiatric nursing homes
 - a religious community or cloister with more than 8 persons
 - a prison
3. People in newly created households that have not yet registered as such in the NR.

4.2. Sampling methodology

4.2.1. Sample Size

For all HIS up to 2013, the total number of successful interviews for the basic sample was set to 10.000 (3500 for Flanders, 3500 for Wallonia, including 300 for East Belgium and 3000 for Brussels). This sample size was based on sample size calculations performed during pre-analyses for the HIS in 1997, taking into account specific budget constraints and the available logistic means. For the HIS 2018 the Commissioners agreed to allocate 700 extra individuals to Flanders. This was to enable the Flemish Region to obtain health-related estimates at the level of the “zorgregios”, even though their precision may not be optimal (see paragraph “motivation” under point 4.2.2). For the HIS 2023 this remains unaltered, resulting again in a basic sample size of 10.700 (4200 for Flanders, 3500 for Wallonia, including 300 for East Belgium and 3000 for Brussels).

On the basis of the previous surveys, the efficiency obtained in estimation at the national and regional level appeared to be sufficient. It is however rather small for estimation purposes at the provincial level, especially for small provinces. Therefore the Protocol Agreement of the commissioners includes that provinces who wish additional sample units are entitled to ask for an oversampling of their province, if they are prepared to provide the financial means for this. For the HIS 2023, as in 2018, no provinces asked for an oversampling. Yet, to obtain more precise estimates for East Belgium, the German Community authorities requested and provided funding for an oversampling of 600 units. As a result, the final sample size for the HIS 2023 is 11.300. This includes the basic sample (10.700) and the oversampling of East Belgium (600).

4.2.2. Stratified Clustered Multi-stage design

Motivation

In the design of the sampling scheme, both the coverage of the population in Belgium and the logistic feasibility of the fieldwork are important concerns. Even when a relatively exhaustive list is available (such as the National Register) a simple random selection from this list would be expensive from a practical point of view because the spread of households to be interviewed would be too wide and scattered. By using a more complex sampling method it is possible to obtain a larger sample size than would be obtained by simple random sampling at the same cost. Therefore, a multiple stage clustered sampling design is preferable. In this design, municipalities serve as primary selection units, while households within the municipalities and individuals within households are respectively second and third-stage units.

Choosing a stratified sample instead of a simple random sample is motivated as follows. Sample surveys displaying small variability among the measurements will produce small bounds on the errors of estimation. In other words, stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size.

This result is enforced if the strata are largely homogeneous. For the HIS 2023, there are three stratification levels, one at the regional, one at the provincial and one at a lower geographical level, which is the “zorgregio” in Flanders and the “arrondissement électoral” in Wallonia. The reason for this third level of stratification is the request of the Flemish Community to be able to make to some extent comparisons between zorgregio’s. In Wallonia no similar geographical entity exists, but as the size and number of arrondissements électoraux in Wallonia is comparable to the size and number of

Flemish zorgregio's, it was decided to take the arrondissements électoraux as third stratification level in Wallonia. Within the Flemish and the Walloon Regions a proportional representation per province in the base sample is sought. A simple random sample of municipalities within a region would ascertain this condition from the sampling framework point of view. Resulting differences are regarded as purely random. However, stratifying proportionally over provinces controlled this random variation further. Similarly within a province the number of units per zorgregio /arrondissement électoral is proportional to the population size, but the minimum number per zorgregio/arrondissement électoral is set to 200.

Municipalities are established administrative units, they are stable (in general those units do not change during the time the survey is conducted), and they are easy to use in comparison with other specialised sources of data related to the survey. Municipalities are preferred to regions or provinces, because the latter are too large and too few. The great variation in the size of the municipalities is controlled for by systematically sampling within a zorgregio/arrondissement électoral with a selection chance proportional to their size.

Within each municipality, a sample of households is drawn so that groups of 50 individuals can be interviewed in total. Clustering also takes place at the household level since members of the same household are more alike than persons not belonging to the same household.

Whereas the stratification and the systematic sampling according to municipalities usually increase the precision, the clustering effect (selecting groups of 50, selecting households instead of individuals) will slightly reduce the precision, since units will resemble each other more than in a simple random sample. However, since stratification is based on unequal probabilities (to guarantee meaningful sample size per stratum), only a slight decrease in overall efficiency is to be expected. The effects due to clustering and stratification observed in the HIS 1997 are very mild and do not outweigh the advantages. This design choice indeed enables persons to be sampled from abbreviated listings and, hence, reduces the survey field worker's travel distance significantly.

In summary, in the light of the previous remarks, multistage sampling is the appropriated way to get access to individuals. An overview of the steps in the selection procedures is given in the next section.

Overview of the design

In summary, the final sampling scheme, i.e. the mechanism to get a probabilistic sample of households and respondents, is a combination of several sampling techniques: stratification, multistage sampling and clustering. The selection process consists of the following steps:

1. Regional stratification. Belgium is divided into 3 regions, the Flemish Region, the Walloon Region and the Brussels Region, for which the number of interviews has been predetermined. The reason for this stratification is to ensure that we can make/draw inferences for each region with nearly the same precision. The number of interviews to be carried out is fixed to 4200 for Flanders, 3500 for Wallonia and 3000 for Brussels. These figures do not include the oversampling.

2. Stratification at the level of the provinces. This second level of stratification is done to improve the quality of the sample over a simple random sample. In particular, a balanced geographical spread is achieved. For the base sample, the sample size within the provincial stratification is proportional to the population size of the province.

3. Stratification at the level of the zorgregio's/arrondissements électoraux. The third level of stratification is done to be able to make geographical comparisons at a lower geographical level (especially comparing zorgregio's in Flanders). The sample size of East Belgium (which is geographically located in the province of Liège) is predetermined by convention. Moreover, this region asked an oversampling for the HIS 2023. Therefore East Belgium is considered as a separate stratum within the arrondissement électoral "Verviers".

4. Then, within the strata, units are accessed in two (for the households (HH)) or three (for the individuals) stages:

(1) Municipalities are selected with a selection probability proportional to their size, within each stratum. These municipalities are called the Primary Sampling Units (PSU). To facilitate the fieldwork, for each PSU selected, a group of 50 individuals residing in that municipality has to be interviewed successfully during the year 2023.

(2) Within each municipality, a sample of households - the Secondary Sampling Units (SSU) - is drawn in such a way that 50 individuals per PSU can be interviewed in total.

(3) Finally, at most four individuals - the Tertiary Sampling Units (TSU) - are chosen for the interviews within each household. Only questioning the reference person within a household would not enable us to give a good picture of a household's health status. For households with four members or less, all members are selected. For households with at least five members, the reference person and his/her partner (if any) are selected. Among the remaining household members a random selection is made, so as to yield four selected household members. Interviewing more than four persons within a household is inefficient because of the familial correlation and because the burden on the household would be too large.

5. To avoid seasonal effects, interviews are spread over the whole calendar year so that each quarter is comparable in terms of number of selected units. The quarters are defined as follows: Q1: January-March; Q2: April-June; Q3: July-September and Q4: October-December.

Detailed information about the sampling procedures can be found in the document "1- sampling2023 reference document"⁴.

4.3. Instruments

4.3.1. Mode of data collection

Data collection in the HIS 2023 takes place using two standardized questionnaires:

1. a questionnaire administered in a face-to-face interview setting, and
2. a paper questionnaire handed out to participants for self-completion.

The face-to-face interview

The questionnaire for leading the interview is intended for all the selected individuals (maximum four) within a household. The data are collected via a "computer assisted personal interview" (CAPI). Here, the interviewers read out the questions on screen to each respondent and enters their answers directly

⁴ Stored on the Sciensano server at: \\sciensano.be\fs\1140_DATA\HIS\HIS2023\Research protocol

into the programme displayed on their laptop. In some specific circumstances, the interviewers can call upon a “proxy interviewing” strategy: this means that a person (belonging or not to the household) is allowed to respond on behalf of the selected participant, because this participant - for a definite reason – is unable to reply her/him-self. By default, children under the age of 15 are not interviewed themselves, but usually via their parents. For one particular topic of the face-to-face questionnaire, i.e. children’s mental health, the interviewers may ask the proxy-parent to fill in the questionnaire directly on his CAPI computer if the parent is willing/able to (this mode of collection is called “CASI”: computer-assisted self-interview”).

The self-administered questionnaire

The questionnaire designed for self-completion collects information from selected participants aged 15 years or older. The interviewees fill in themselves the paper-and-pencil questionnaire that is handed out to them, without the intervention of the interviewer. The decision to use such a mode of data collection is based on the idea that some topics, such as mental health, alcohol consumption, drug use, and sexual health for instance, are quite sensitive and are therefore not suited for a face-to-face interview, because the replies could be influenced by social pressure/desirability. The self-administered questionnaire cannot be completed by a proxy, by rule.

The questionnaires are available in French, Dutch, German (the three national languages in Belgium) and English (for foreigners). If a selected individual does not speak any of these languages, a translator can be used. The translator is generally a member of the household who knows both the kin language and one of the languages spoken in Belgium, mostly through schooling. Translating, as in this case, is not considered as a proxy interviewing because the selected person still replies for her/him-self, and the translator does not reply on her/his behalf (no ‘filter’).

4.3.2. Conceptual development of the HIS questionnaires

The guiding principle of the questionnaires’ development is to keep comparable questions across surveys to allow the study of time trends. This does not mean that the questionnaires are static. In fact, the content of the questionnaires evolves over time to address emerging concerns in the arena of public health. When new needs are identified, the conceptual development of any new set of questions starts with an extensive literature review to identify the best indicators that describe the issue to be monitored. The identification of these indicators depends on the availability of instruments that are valid and reliable or instruments that have previously been used in other high quality surveys.

For the HIS 2023, the questions have been developed based on the following guidelines:

- [*Previous HIS questions*](#): The questions proposed in the HIS 2023 are grounded on the questions available in previous HIS editions to allow comparison across surveys and monitor the evolution of health indicators in time.
- [*Commissioners’ demands*](#): New “modules” (= set of questions related to specific topics) are requested as new health concerns emerge in the society. Such requests may be issued by the commissioners of the HIS and by external users of HIS data, such as health agencies or academics. Requests by the commissioners are given higher priority.
- [*Requests for international instances*](#): International organisations such as the WHO, EMCDDA, OECD or UNAIDS claim periodic delivery of a set of health indicators. These indicators need to be included in the HIS in order for Belgium to fill such an international requirement.

- [European Health Interview Survey](#): For the HIS 2023, there is no need to comply with the request of questions/indicators from Eurostat in the context of the European Health Interview Survey (EHIS) as was the case for the HIS 2018. This is because the next EHIS is scheduled to take place in 2025 and Belgium did not receive a derogation to organize it in 2023. For EHIS, a specific (on-line) survey will take place in 2025.
- [Expert consultation](#): The questions proposed for some modules are discussed in working sessions with scientific and academic experts, members of health agencies or administrations, and fieldwork experts. Some are reviewed by experts in the context of an e-mail consultation.
- [Approval of questionnaire](#): The questionnaire must be submitted to, and approved by, the commissioners' committee, all ministerial cabinets and the inter-ministerial conference on health.

In addition, the development of the HIS questionnaires is guided by the need to keep an optimal length, meaning all demands for additional topics/questions cannot be honoured. The questionnaire used for the previous HIS took on average one hour per person to complete. Therefore, any new question to be included in the HIS 2023 implies dropping other questions. This approach must be taken in order not to jeopardize the participation to the study and the validity of the results.

Finally, the decision process that leads to the development of the HIS questionnaire is specific to each module. A "conceptual paper" is developed to document this process for every module included in the HIS. These documents are archived as: XX_2023_cpt.doc (XX standing for the two letters of identification of each specific module) to be found in: \HIS\HIS2023\modules\XX. The conceptual papers follow the same structure and generally include items such as the place of the module in the conceptual framework of the health survey, the definition of concepts and their relation to health, their implication at the level of the society, the added value of having the module in the HIS, the relative importance of the information for the decision makers, other potential source of similar data in Belgium, a review of available instruments, the choice and reason for opting for a specific instrument to include in the BHIS, etc.

4.3.3. Content of the questionnaires

The HIS questionnaires examine topics that can be grouped in five broad health domains (described in section 3.2. section) together with a series of background socio-economic variables. The information gathered on these topics are organised in what is commonly called "modules". The list of modules that are under scrutiny in the HIS 2023 is shown in the table below. The modules either address the household as a whole or address the individual members of the household.

Household questionnaire

Certain modules collect general information from the reference person that is relevant for all household members, so they are only asked once per household. These modules are relative to the household composition, ethnicity, household income, burden of health care expenses, exposure to tobacco smoke or to e-vapour inside the house, housing conditions and nuisances.

Individual questionnaire

All the other modules collect information directly from the selected household members (whether by PAPI or CAPI/CASI mode), though some modules are limited to specific age groups or gender.

Questionnaire development procedure

The questions included in the HIS 2023 as well as in the previous surveys are developed in the wake of the conceptual papers and stored apart in an Excel file that is elaborated for each module. This excel file allows the comparison of the questions across all survey years and all four languages. The documents are archived on the Sciensano server under the name: "Overview_questions_XX" placed in the directory \HIS\HIS2023\modules\XX (where XX stands for the two letters of identification of each specific module).

The questionnaires of the HIS 2023 will be made available on the HIS website "www.healthsurvey.be".

TO FINALIZE when decisions are made

MODULES	Abbreviation	Face	Self
<i>Background information</i>			
1. Information on the selected individuals	NR	x	
2. Household composition + ethnicity	HC	x	
3. Education	ET	x	
4. Income	IN	x	
5. Employment	EM	x	
6. Absence from work	AW	x	
<i>Health and well-being</i>			
7. Subjective health	SH		x
8. Illness and chronic conditions	MB	x	
9. Specific diseases and conditions	MA	x	
10. COVID-19	CO	x	
11. Long-term limitations	IL	x	
12. Dental health / Dental care	DH / DC	x	
13. Frailty	FR	x	
14. Bodily pain	PI	x	
15. Mental health:			
– Stress and well-being	WB		
– Energy/vitality	VT		x
– Depression and anxiety disorders	AD		x
– Suicidal behaviours	SU		x
– Eating disorders	EB		x
– Problematic gambling	GA		x
– Ecological anxiety	EA		x
MODULES	Abbreviation	Face	Self
16. Child & adolescent mental health	CH	x	
17. Quality of life	QA	x	
<i>Health behaviour and lifestyle</i>			

18. Alcohol consumption	AL	x
19. Tobacco smoking	TA	x
20. E-cigarettes use	EC	x
21. Use of illicit drugs	ID	x
22. Physical activity	PA	x
23. Nutritional status	NS	x
24. Nutritional habits	NH	x
25. Sexual health	RH	x
26. Screen use		
Prevention		
27. Cancer screening	SC	x
28. Vaccination	VA	x
29. Health literacy	HL	x
Health care consumption		
30. Contacts with the general practitioner	GP	x
31. Contacts with the specialist	SP	x
32. Contacts with other health services:		
- Paramedics	PM	x
- Non-conventional medicine	NC	x
- Home care services	HS	x
33. Contacts with the dentist	DE	x
34. Use of medicines	DR	x
35. Accessibility to health care	AC	x
36. Patient experiences	PE	x
Health and society		
37. Lodging conditions	LO	x
38. Health and environment	HE	x
39. Violence	VI	x
40. Social health	SO	x
41. Informal care	IC	x

For the 2023 edition, two collateral health data sources (dental examination data collection and AIM/IMA - records) will be linked to the HIS 2023 dataset, offering some “objective” information that was previously obtained by means of interviewing the participants. This implies that the HIS 2018 questionnaire could be shortened, allowing space for new research topics. Because these linkage projects have protocols of their own (which will be referenced at a later stage). The aim is merely here to mention their existence and content:

Linkage to the dental examination data

A subsample of the HIS participants are re-contacted on a voluntary basis for a dental examination that will provide standard information on parameters such as dento-facial anomalies, oral hygiene status

(plaque index), periodontal status (DPSI index), tooth wear status (BEWE index), tooth status (DMFT index). The full list of collected parameters is found in the oral health survey protocol.

Linkage to IMA data

The reimbursement of expenses of patient's health care (ambulatory health care providers, hospitalisation, emergency department, dentist, psychiatric care, medicines) and the corresponding code system are stored in a yearly database that will be linked to the HIS 2023 data.

4.4. Fieldwork procedures

4.4.1. Interviewers' tasks

To collect the data through the HIS 2023 questionnaires, more or less 200 interviewers are needed. The basic role of these interviewers is twofold: (1) to establish a contact with all households selected for interview in the group assigned to them, and (2) to conduct the interviews with the selected members of the participating households. One should bear in mind that the burden for the interviewers (in terms of time spent) is mostly related to the efforts to establish a contact with a selected household. The procedures foreseen for these tasks are the following:

Establishing contact with the households

As described further on, interviewers receive the addresses of the households they have to contact, together with the names of the members of the households. No telephone or mobile number nor e-mail address is provided. It is up to the interviewers to try to contact the "activated" (invited) households either by telephone or at doorstep. The general rule is that the interviewer has to try to contact the invited household at least 5 times. Every contact attempt has to be documented, using a contact sheet.

In some rare cases, the information derived from the National Register (sampling frame) can be incorrect, that is:

- The address of the household provided to the interviewer does not exist
- The address of the household provided to the interviewer does exist, but it is clear that the household does not live there (e.g. the building is a ruin, the building is clearly abandoned,...)

In such cases, interviewers can indicate this info to the contact sheet and can stop their contact attempts.

In other cases, it is clear that the selected household is not eligible for participating in the survey:

- The address of the household is the address of a prison, a psychiatric institution or a large monastery

Also in such cases, interviewers can indicate this info to the contact sheet and can stop their contact attempts.

In case (a) the address of the household exists, (b) there are no signs that the household does not live at the indicated address and (c) the household is eligible for the survey, interviewers can label the household only as 'non-contactable', 'refusing' or 'participating' household:

- If a household cannot be contacted despite 5 (duly documented) attempts, the household can be labelled as 'non-contactable'.
- If a contact with the household takes place, the interviewers have to ask the household (reference person or his/her partner) if the household is willing to participate in the survey.
 - If the household does not agree to participate, the household receives a "refusal" status.
 - If the household agrees to participate, an appointment should be made (date and time) to conduct the interview(-s) in that particular household.

Conducting the interviews

The structured interviews are carried out in a face-to-face setting using the CAPI mode of data collection with all selected household members (or proxy interview for those under 15 years of age). The data collection also consists of handing out a self-administered questionnaire to the selected members 15 years or more (PAPI mode). The procedures related to the HIS data collection are documented in the "interviewers' manual" and explained during the interviewers' training.

4.4.2. Fieldwork paradata

Launching a large-scale survey such as the HIS implies dealing with numerous data at various levels, namely, data stemming from fieldwork operations (also referred to as "paradata") and data collected from the HIS participants (resulting from applying the HIS questionnaires). In the process of collecting survey-based data, fieldwork procedures are foreseen which generate the need for tight management by means of specific follow-up indicators.

Basically, the goal of the data collection phase is to obtain the participation of as many households needed so to reach a total of 11.300 interviews (baseline sample of 10.700 and supplementary sample of 600 interviews in Eastern Belgium) during the calendar year 2023. Because the HIS is not a compulsory survey, a certain number of selected households will not participate. Non-participation (including non-eligible households, non-contactable households and refusing households) is an important issue in the data collection phase, especially since a sophisticated substitution process of non-participating households is applied in the HIS.

The data collection phase starts with the "activation" of the households, that is, when they are notified that they were randomly selected to take part in the HIS 2023. In order to inform a household that it is selected for participation in the survey, SB sends an invitation letter and an introductory leaflet addressed to the reference person of the household. The leaflet explains, amongst others, the goal and the content of the HIS as well as the contact practicalities, i.e., that an interviewer will contact them shortly after receiving the letter. In addition, there will be a video to better explain to the participants the interview process. Households are explicitly informed that participation is not compulsory and different means are offered for them to express their refusal of further contact (e-mail address, free-telephone number, coordinates of the HIS team members). In the letter, reference is made to the possibility to participate in the Oral health survey.

As soon as these information documents are sent to the household, the interviewer responsible for the given (group of) households receives access to the contact data of the household (full name reference person, full address, first names of the (other) household-members), with an indication which household members are selected for participation in the HIS2023 and can start attempting to establish a contact to find out whether the household is willing to participate, and if so, to set an appointment for the interview phase. This so-called contact information is retrieved from the National

Register used as a sampling frame for the selection of the households and a selection of the household members invited for participation.

It is important, for both the interviewer's work as well as for a close monitoring of the fieldwork, that all attempts to contact the household are duly documented. For this purpose, the interviewer has to complete a (computerised) "contact sheet" for every activated household in his/her contact list. This contact sheet is a (slightly adapted) standard sheet used by SB in their own surveys.

For every attempt and/or successful contact with an activated household, the contact sheet must be updated with the following data:

- The date
- The day
- The moment of the day of the contact
- The mode (at doorstep / by telephone)
- The outcome of the contact:
 - Interview conducted
 - Interview scheduled (+ date scheduled for the interview)
 - Refusal (*)
 - Interview impossible (*)
 - Not available on the moment of the contact
 - Nobody opened the door (*)
 - Nobody at home (*)

For results marked with (*) the interviewer has to describe – if possible – the situation in detail.

Because of the substitution process implemented for non-participating households in the HIS 2023, the contact sheets must be completed on a regular basis, preferably day-by-day. Since the communication system between the interviewer and SB is web-based, daily updates will enable SB to monitor the data-collection proceedings 'in real time'. That way, special attention can be paid to households 'on hold' (meaning that their final outcome status is not yet known). In case a household is still 'on hold' after 6 weeks of activation, the responsible interviewer will be notified that he/she must do the maximum to assure that this household gets its final status.

As stated previously, a substitution process is applied to the non-participating households: the non-participating household are replaced with a predefined household (the next one in the sample cluster). Practically, when SB is informed that an invited household will not participate, an introductory leaflet and invitation letter is sent to the substitute-household, with a subsequent notification to the interviewer that contact attempts can start with the replacement household, for which s/he receives an open access to the contact data.

If the household consents to participate, the individual data collection through CAPI can take place. The CAPI application indicates how many (max. 4) and which household members have participated. This information is not readily available regarding the self-administered questionnaire (PAPI). It is the interviewer who must mention the number of household members that completed the self-administered questionnaire. These completed questionnaires should be sent by the interviewer to SB. Until SB receives these questionnaires, they fall under the responsibility of the interviewer. An interview with a household member of 15+ for which no proxy was used, consist of two parts:

information collected through CAPI and the self-administered questionnaire. Before considering an interview as executed, both parts should be available at SB.

4.4.3. Pool of interviewers

To reach the target of 11.300 interviews (divided in 226 groups of 50 interviews) in one calendar year, the HIS requires the services of about 200 interviewers, depending on the number of groups some of them take in charge.

Recruitment

Interviewers that carry out the HIS 2023 are selected from a pool of interviewers already active in other surveys of SB. Given the specificity of the HIS – which is conducted in selected target municipalities – extra interviewers will have to be recruited. The recruitment of candidate-interviewers is a joint Sciensano – SB initiative: in case no interviewers are directly available in the pool, local administrations and state-run institutions (post points, schools, hospitals ...) will be contacted to search for candidates.

Training

All interviewers active in the HIS 2023 have to follow a collective one-day training in order to ensure the standardisation of all data collection and fieldwork procedures. Three major themes are addressed during the training session:

1. The overall Statbel approach applied in data-collection (the use of CAPI, communication with SB, contact sheets,...);
2. Specificities of the HIS 2023 survey (content of the questionnaire, conceptual background, selection of household members,...);
3. Supplementary tasks in the context of the Oral Health Survey (introducing the survey, “intent” form,...).

Interviewers’ training is not only organised in the preparation phase of the HIS 2023 (i.e., late 2022), but also on a regular basis during the fieldwork in 2023 because of interviewer drop-out and turn-over throughout the data-collection phase. Of course, all possible efforts are made to limit drop-out. In this sense, the candidate-interviewers are fully informed before start of the actual workload and efforts expected of them, the kind of problems they will face during data collection (e.g. when contacting households) and how much they are paid. Interviewers active in the HIS 2023 work under the responsibility of Statbel; their tasks and their wage (expressed in terms of conducted interviews) will be published as a Royal Decree in the Official State Bulletin.

Guideline manuals⁵

The procedural guidelines for the interviewers are provided in three manuals:

- 1) [*Contact-procedures*](#) (Manual 1). An important feature of the survey is that the work of the interviewer is standardised as maximum as possible. This is not only related to the moment of the interview itself, but also for what concerns the contact-procedures (see data management procedure).

⁵ Stored on the Sciensano server at : \\sciensano.be\fs\1140_DATA\HIS\HIS2023\Interviewers\Manual

- 2) [Content of the questionnaires](#) (CAPI & SELF-ADMINISTERED) (Manual 2). Although the bulk of the work of interviewers consists in running through a CAPI questionnaire, it is important that they have an insight in the aim of the questions, the meaning of the concepts, etc. This information is of interest for the interviewer in case the respondent wants to have extra-information on a specific question or topic. In this context, an overview is also provided on the content of the self-administered questionnaire; although interviewers are not expected to run through this questionnaires (respondents complete the questionnaire themselves), it is found useful that interviewers have at least some notions of its content. The last part of the interviewers' task is to introduce the dental health survey as well as future linkages and follow-up surveys. In the manual, a short overview is provided on these points.
- 3) [Instructions for the CAPI application](#) (Manual 3). Some interviewers (working for other surveys of SB) have already some knowledge and experience of working with CAPI, but for others, the HIS 2023 will be the first survey in which this tool is used. In this context, it is important that clear and thorough instructions exist for what the use of CAPI concerns. While this theme is addressed during the training session, it remains important that interviewers have a manual at hand in case they need clarification or are confronted with a technical problem.

Newsletters

Whenever systematic errors in the conduct of the interviews are reported, or FAQ come up, reminders and extra explanation will be given to the interviewers in the form of ad hoc newsletters. These contact letters may also contain tips, progress facts and other anecdotal information to motivate and sustain the fieldwork rhythm.

4.4.4. Fieldwork control

Using an online contact sheet enables SB to closely monitor the fieldwork, with specific SAS® - programs and fieldwork indicators developed for this purpose:

- Given the overall target of the data-collection phase, having 11.300 completed interviews by the end of 2023 (with prefixed numbers of interviews in each of the regions, and the prefixed number of interviews to be realised in East Belgium), a follow-up of the accrual-rate is necessary. Hypothetical accrual projections, combined with comparative real accrual rates (in each of the regions and provinces), enables to forecast if the target(s) will be reached on time. In case of negative forecasts, appropriate measures can be taken to optimize the realization of the target(s). These measures depend on the source of the problem, but they usually concern discharging interviewers and/or hiring new ones, giving extra pay to interviewers for the regions where participation is more difficult to obtain, etc.
- A follow-up of the accrual-rate at the level of each particular group (of 50 interviews assigned to an interviewer) allows tracing outlier-interviewers (in both positive as negative terms). For the positive outliers (interviewers with an accrual rate far quicker than that of others), a detailed assessment of their activities based on information derived from the contact sheets should be undertaken to estimate whether their work complies with the expected procedure. Negative outliers (interviewers with a substantial lower accrual rate) are contacted by phone to verify why they have a distinct profile. If necessary, Sciensano and SB will jointly decide to replace the interviewers that do not work in line with the prescribed procedures. Based on a

thorough check of their work, including re-contacting the households for further information, a decision will be taken as to keep or discard the interviewer's data collected so far.

- Data collected in the context of the Oral health survey (that is: households that have indicated to agree with a re-contact in the context of the oral survey) needs close follow-up.
- Next to a rather quantitative data monitoring, the information present in the contacts-database should enable to have an insight in the quality of the interviewers' work. Again, the basic question here is to check to what extent the procedure interviewers apply is in line with the procedure they should apply. Specific indicators are developed to monitor the contact attempts to trace interviewers who do not perform the fieldwork as prescribed.
- A supplementary quality control is carried out by means of a short standard postal survey addressed to all participating households. In this survey, specific questions deal with the work of the interviewer: did the interviewer respect the protocol for data collection (home based face to face interview), did he/she explain the objectives of the study and were they clear enough, was he/she pleasant, polite, clear, patient, professional...? If deviations are reported in various households, it can be decided to reject all the completed interviews and dismiss that interviewer.

4.5. Data flow and data management

4.5.1. Data entry procedures

There are two modes of data entry for the HIS 2023 data:

- Data from the face-to-face questionnaire that are collected using the computer assisted personal interviewing (CAPI) approach do not need a specific data entry procedure as the data are entered directly during the interview. After conducting the interviews, the data collected are directly uploaded in the HIS 2023 database at Statbel without further manipulation.
- Data from the self-administered questionnaire (PAPI data collection mode) require manual data-entry. For this, the interviewers are required to send the completed self-administered questionnaires to Statbel for transcription, where a team of professional typists is responsible for the data-entry. Sciensano has developed a specific data input programme in Blaise® application. The team of typists has to follow a short training session during which the questionnaire and the application are presented. A manual for data-entry operations is provided by Sciensano to ensure standard transcription. The data entry process in Blaise® must start shortly after the beginning of data collection.

4.5.2. Data transfer from Statbel to Sciensano

Data for every participating and non-participating household are transferred from Statbel to Sciensano when all the expected information for a certain household has been uploaded in the database, both directly from the CAPI application and the PAPI application after data-entry. Based on the information on the contact form and the CAPI application, Statbel can verify if the declared number of self-completed questionnaires has been subject to data-entry.

The data transferred from Statbel to Sciensano include:

- the case identifier (i.e., HIS-specific code for the household and the household members);
- the follow-up para-data (derived from the contact-form) in order to study the quality of the data-collection and to study non-response;
- the data about age, sex, nationality and household composition of non-participating households;
- the survey data from the CAPI and PAPI interviews.

Data-transfer from Statbel to Sciensano is enabled via secured FTP. A first data set is transferred to Sciensano at the latest on March 1st 2023 in order to check its quality and consistency. The final dataset is transferred to the Sciensano at the latest on March 1st 2024.

4.5.3. Record identifier

To identify respondents in the context of Statbel-Sciensano communication, a system of a unique HIS case identifier for each respondent is devised. A unique case identifier code consists of the identification code of the household (10 digits) plus a 2-digits code that refers to the place of the individual within the household. The household code is based on the municipality where the household resides (4-digit number), the trimester when the household was selected (1-digit number, from nr 1 to nr 4), the cluster place (2 digits: the cluster number) and the HH rank number within the cluster (3 digits).

A key data set comprising both the unique HIS case identifier at the individual level and the individual code number that was produced by Statbel with an algorithm applied to the National Number of selected households, is kept by Statbel.

4.5.4. Creation of a working database

After the data is exported to Sciensano, program files are created in SAS to generate an initial working database. The programmes deal with:

- input of the data (CAPI – PAPI – contact information)
- monitoring of the field work
- checks of consistency
- creation of 3 SAS data files: one that includes information on the household level, another that includes information at the individual level and one that includes information for each medicine that was taken in the 24 hours before the interview.
- allocation of labels in SAS in 3 languages (Dutch, French, English); labels in English are also available in a separate excel file
- creation and/or cleaning of general background variables:
 - age, sex;
 - demographic information: nationality, country of birth, region and province of residence, urbanisation level of municipality of residence, household type;
 - socioeconomic information: income, education, occupation;
 - housing situation.
- creation of survey weights at the level of the individual and household

All the procedures performed are documented within the programme itself, following specific Sciensano's standard operational procedures (SOP).

4.5.5. Controls of coherence

Checks performed during data entry

The Blaise® data-entry programme (used for entering the self-completed questionnaire) and the CAPI programme used during data collection (face-to-face questionnaire) contain several mechanisms aimed at the production of quality data. These controls include the following:

- The system follows the logic of the questionnaire. When a question is not applicable, the data entry programme jumps to the next relevant question.
- Some questions only need one answer (only one field for the question in the table); others allow several answers (one field per answer in the table).
- Some answers are accompanied by a free text; data entry in this field is only allowed if answer to previous question meets specific criteria.
- The system only offers a limited number of possible answers. If one tries to insert another value, an alert shows up.
- Several answers have a table's look-up to choose the right information in a list instead of typing it out: countries, professional activities and medicines.

Prior to the actual data entry, the programme is tested and adapted by members of the HIS team by using mock-up questionnaires.

Checks performed on the working dataset

After the creation of a working database, two types of data checks are performed to create a set of final clean databases:

1. Vertical checks

Vertical controls are intended to verify for each household and for each individual within the household whether the information and records are conform to what is expected. Mainly, it is important to check whether each person who has participated in the survey based on the contact form and the face-to-face questionnaire has also completed a self-completed questionnaire.

If inconsistencies are noticed, they are verified (e.g. by physical checks of the autoquestionnaire by the secretariat of Statbel) and if needed, corrections are made in the input programme. The corrections are documented in the programme.

2. Horizontal checks

Next to the vertical checks, it is necessary to control for internal consistency of the data. Horizontal controls are intended to verify whether the answer to a specific question is coherent with the rest of the questionnaire. Thanks to the utilisation of CAPI, routing errors and inconsistencies can be avoided as the system is programmed to guide the interviewer through questionnaire. Therefore, the data collected via CAPI will undertake no intensive data checks. Depending on the module some conceptual editing may be needed.

Data collected using the PAPI questionnaires require a specific phase of data checks because in this case errors may have occurred during data entry. In this case, checks for routing errors as well as for inconsistent replies will be undertaken. These controls are performed separately for each module of the PAPI questionnaire.

The procedures for data checks are documented in the SAS programmes archived as: Check_XX_2023.sas (XX for the two letters of identification of each specific module).

The errors identified through these controls are pooled together in one document and sent back to Statbel in order to have verifications performed (i.e. go back to the paper questionnaires to check if reply is correctly encoded). Corrections that are applied are sent back to Sciensano in a new version of the data file, exported from Blaise®. When corrections cannot take place (meaning that the error cannot be corrected without re-interviewing the individual) the incoherencies are taken into account in the programmes developed for the creation of the indicators.

4.5.6. Storing and archiving

A number of databases are to be created for different purposes. A description of these databases is provided below. These databases are stored on a database server at Sciensano. A backup of the databases is done every day on a separated hard disk, and this file is backed up and archived as another file.

Final databases

A final common database is to be created in which the variables and indicators for all the modules are available and in which data from the previous surveys are included, as far as they are also available in the HIS 2023. The procedures to be followed for the creation of this database are included in an automated programme “01_Stg_all_tfm_2023.sas” and “01_DWH_all_2023.sas” in the directory \HIS\HIS Data Warehouse\. The procedures are documented within the programmes.

As a result, 3 final databases are created:

- his2023_hh_vxxxx : information at the household level
- his2023_ind_vxxxx: information at the individual level
- his2023_dr_vxxxx: information for each medicine that was taken in the 24 hrs. before the interview

Databases for external users

From the final databases mentioned above, anonymised datasets are created for external users. In these datasets some variables are removed (e.g. statistical sector, date of birth) or aggregated (e.g. age). In the external datasets, the HIS ID's are replaced with arbitrary codes that are different for each external user.

The procedures to be followed for the creation of these datasets are included in an automated programme “01_DWH_all_HIS2023” in the directory \HIS\ HIS Data Warehouse\03_DWH. The procedures are also documented within the programme.

The programme will create SAS data files. A codebook in excel includes all variable and value labels in English. A manual for external users provides all information that is needed to understand and use the data.

All files and documents are saved in the directory HIS\HIS Data Warehouse\03_DWH\

The modalities of the transfer of the data to external parties is described in a document stored in the directory \HIS\HIS2023\external users\

Databases for internet based analyses

From the final databases mentioned above, datasets are created to be used as a source dataset for the Internet based application that allows to carry out interactive analyses via the HIS website.

The procedures to be followed for the creation of this database are included in an automated programme in the directory \\sciensano.be\fs\1140_DATA\HIS\HIS Data Warehouse\04_Datamart\2023: hisia.sas. The procedures are documented within the programme itself.

4.6. Data analysis

4.6.1. Indicator development

Data analysis begins with data cleaning and the construction of new variables called “indicators”. These are defined as “variables that indicate certain conditions of interest” and will be used for the final analysis of the data. In some cases indicators are just copies of existing variables (after cleaning) while in other cases they result from the recoding of a/several variable/s.

The creation of the indicators is performed in a separate SAS programme for each module of the questionnaire. The SAS programme is called: XX2023.sas to be found in \\sciensano.be\fs\1140_DATA\HIS\HIS Data Warehouse\02_STAGING\01_Survey\2023.

All the procedures performed are documented within the SAS programme itself. In addition, when the same indicator is available from previous HISs, these data are added in order to be able to compare the results between the surveyed years. The procedures are included in the same SAS programme and documented within the programme itself.

4.6.2. Plan of analysis

For each outcome indicator crude and age and/or gender standardised rates are calculated. The crude results reflect the true prevalence within a population group. The information is purely descriptive and should be considered with caution when comparing the data between sub-populations. The standardised results are based on a mathematical standardization through a linear or logistic regression and allow comparing the indicators between selected background variables while adjusting for age and/or gender.

Around the crude and the standardised results 95% confidence intervals (CI) are calculated.

The crude and standardised rates are presented in function of a selected number of background indicators: gender, age group, education level, region of residence, year of survey. Depending on the outcome, results are expressed as a proportion, a distribution, a mean, a median or other percentiles. The results are presented in a final report including tables, graphs and an explanatory text. They are reported for the whole of Belgium, but also separately for the three regions of the country.

Further exploration of the data varies from one module to another. When relevant, extra analyses are performed including also other variables.

4.6.3. Software selection

Analyses are carried out using SAS version 9.4

4.6.4. Treatment of missing values

In the database three types of missing values are distinguished: 1) not applicable (if the question is not supposed to be answered); 2) no answer (no information available); and 3) “does not know” (if this option is actively indicated in the questionnaire). For each variable and indicator, details about the codes of the missing values are found in the codebook.

Missing values are not considered in the analysis, meaning complete case analysis is carried out on each individual indicator (not that each individual must have complete records).

4.6.5. Programming

A number of programs are developed for the data management and the analyses:

1. the program to construct a cleaned global dataset (the raw dataset is received from SB); this program includes also the construction of the survey weights;
2. the programs dealing with the analyses of the methodological aspects of the HIS (description of population, participation rate, etc.);
3. the programs dealing with data cleaning and construction of indicators for the individual modules;
4. the programs to produce crude and adjusted results, trend analyses and more advanced analyses
5. the programs constructing the final databases used for the interactive analyses via the website and the final databases for external users

Program to construct the cleaned global dataset (including weights)

A program in SAS with as input the data from the Blaise® data entry programme (that are received from SB) produces the cleaned global HIS 2023 dataset that consists of 3 databases: one database with data at the individual level, one database with data at the household level, one database with data on the medicines that are consumed in the past 24 hours.

The program consists of the following steps:

- Data input from SB files
- Construction of new variables at household level (e.g. region, province, degree of urbanisation, district, status of household, education at household level, equivalent income of household)
- Construction of new variables at individual level (e.g. age groups, living condition of elderly, nationality)
- Correction of inconsistencies
- Calculation of survey weights

A technical document describing the calculation of the survey weights in the HIS 2023 is available on the server⁶ and will be placed in the manual for external users of the HIS 2023 micro-data.

Programs dealing with analyses related to the methodological aspects

A set of programs deals with methodological aspects of the HIS such as:

⁶ Stored on the Sciensano server at: \\sciensano.be\fs\1140_DATA\HIS\HIS2023\Research protocol

- Calculation of time needed to complete the oral questionnaire
- Reason for use of a proxy
- Homogeneity of the households belonging to the same cluster
- Description of households in function of their participation status
- Reasons for non-participation
- Description of the sample in function of background characteristics

Programs dealing with data cleaning and construction of indicators for individual modules

These analyses are done by module. For each module the program consists of the same steps:

- Input of relevant variables of HIS 2023 database
- Input of comparable data previous HIS cycles.
- Data cleaning (correction of inconsistent data), allocation of missing values, formats
- Calculation of the indicators as described in the concept paper
- Computation of database to be used as input file for the basic tables (see next point)
- Computation of basic tables with the crude and adjusted results by background characteristics
- Computation of data files to be used as input for the program constructing the database for external users.

Programs to produce crude and adjusted results, trend analyses and more advanced analyses

Macros are developed for the following types of analyses or outputs:

- Basic tables for the final report with crude and adjusted results (+ 95% confidence intervals) for percentages and means in function of background characteristics
- Graphs presenting percentages or means by age and sex with 95% confidence intervals for the final report
- Trend analyses for Belgium and each of the regions
- Analysis of social inequalities in health

Programs constructing the final databases used for the interactive analyses via the website and the final databases for external users

Input files for those programs include both the cleaned global database and the databases with indicators that are produced for each module.

4.6.6. Non-response analysis

Non-response analysis is performed based on the variables received from the National Register and that are therefore available for all contacted household. These variables include age, sex, household composition, nationality, and place of residence.

4.6.7. Inference methods

The estimation of the population parameters mentioned above and their associated variances is based on assumptions about the characteristics of the distribution of the observations in the HIS sample. These assumptions are that the observations were selected independently and that each observation has the same probability of being selected. However the HIS violates both assumptions as it uses a stratified multistage clustered sampling procedure. For logistical reasons the selected households are clustered geographically (per municipality), and also within the household a sub-sample is taken. As a result sample, units are not selected independently, nor are their responses likely to be independently

distributed. Additionally, we are dealing with unequal selection probabilities because of the regional stratification.

To obtain representative results, both at the national and the regional levels, it is necessary to account for the complex sampling design. Correct estimates and valid inferences can be obtained by re-weighting the data, inversely proportional to the selection probability. A weighting factor is calculated⁷ that combines different aspects of the study design into a single, compound weight, thereby simplifying the computations.

In addition to the weights, the inference procedures should also account for the clustering and stratification of the sampling procedure in order to obtain exact variances and standard errors (and thus also confidence intervals). Clustering will decrease the precision of the estimates; hence yield into bigger confidence intervals. Stratification has in most cases an opposite effect: standard errors and confidence intervals become usually somewhat smaller by taking into account the stratification of the sample.

Internal and external validation

Statistical experts are consulted to ensure the validity of the methods, calculation of weights, etc. The programmes for both the data cleaning and the analysis are created by the HIS researcher responsible for the given module and systematically verified through internal peer review. Internal and external quality checks are also performed for the macros that are produced for standard analyses.

Presentation of results

The final report with tables presenting crude and adjusted results for all indicators are available in pdf-version on the HIS website. Results are presented for Belgium and for each of the three Belgian regions separately.

Via HISIA, the interactive web based analysis tool, it is possible to generate tables with crude results (and 95% CI) in function of background variables that can be selected by the user him/herself (see section 4.10.).

4.7. Methodology used to improve the quality of the data (quality controls)

As stated in the course of this protocol, various safeguards are implemented at different stages of the HIS 2023. The quality assurance procedures include “preventive actions” such as guideline manuals (i.e. for data collection and data entry, for data analysis by external users,...), training (for interviewers and data entry operators) and testing, as well as “control actions”, such as consistency checks implemented in the computer-based programs (CAPI) used during the face to face questionnaire and for data entry, metadata analysis for the fieldwork management, etc. At the level of data collection and management, the following quality assurance procedures are carried out:

⁷ Stored on the Sciensano server at: \\sciensano.be\fs\1140_DATA\HIS\HIS2023\Research protocol

4.7.1. Quality of questionnaires

- **Content:** The questions and questionnaires intended for each particular module of the HIS are discussed with scientific and academic experts, members of the health agencies and administrations and fieldwork experts.
- **Translations:** Questionnaires and their translation (4 languages) are double-checked by the native-language researchers of the HIS team.
- **Pre-test:** The HIS 2023 questionnaires (CAPI and PAPI) are pre-tested in a small but diversified sample of people (gender, social background, age) by the HIS team members themselves. Different features are evaluated, such as the length of the questionnaire and the time to fill it in, the comprehension and the readability of questions, the completeness of response categories, the pathway and skips through the questionnaire, etc. Adaptations are made according to results.

4.7.2. Quality of Interviewing

- **Training:** the candidate-interviewers from Statbel join a full-day training session that includes both a theoretical part (background objectives, fieldwork procedures, content of the questionnaires) and a practical part (on the personal computers provided by statbel).
- **Manual⁸:** the interviewers are provided detailed guidelines containing all aspects necessary to execute the survey correctly. The manuals contain instructions regarding: (1) the contact and fieldwork procedures; (2) the content of the questionnaires; (3) the use of the CAPI application.
- **Fieldwork monitoring:** First, the secretariat established at Statbel supervises the work procedures of the interviewers in terms of schedule and rules for the "contact forms", number of households contacted or interviewed, etc. In case a problem is detected, the secretariat contacts the interviewer to see how s/he may solve it or if it is needed to replace the interviewer. Statbel also assures a helpdesk function for the interviewers. Second, weekly accounts of the number of completed interviews and refusals per region, strata, etc. are sent to Sciensano to make sure the progress of the survey meets the objective by the end of a 12 months period. Actions are taken whenever problems are detected.
- **Questionnaire checks:** The interviewers send the completed questionnaires to Statbel where a first check is performed at reception. The employees verify the number of questionnaires received (regarding the expected number of questionnaires (both F-to-F and auto-Q), as noted on the contact form) and the code numbers of the household members.
- **Control of interviewers:** a quality control questionnaire regarding the work of the interviewers is addressed to all participating households. The questionnaire is returned to Statbel with a prepaid envelope. Each incoming survey is checked for complaints or non-conformity and encoded in an Excel-database. If a problem is identified, the interviewer is contacted to talk it over. In function of the problem, either the right procedure is re-explained, either - if justified - s/he is dismissed from the HIS.

⁸ Stored on Sciensano server at: X:\HIS\HIS2023\Interviewers\Manual

4.7.3. Quality of data entry

- [Manual](#): Regarding the PAPI data entry, guidelines are available for the data entry operators.
- [Program controls](#): Control procedures are built in both the CAPI application for data collection and the data entry program developed for the PAPI questionnaire. They allow avoiding a series of inconsistencies. For instance, the program is constructed in such a way that the data entry is 'guided': every question has a variable field and every variable for which a value should be entered is set with an upper and a lower limit value, making it impossible to introduce values that exceed the specified range. The data entry program blocks or masks entry fields on the basis of information obtained from linked variables (e.g. jumps due to gender: it is impossible to introduce data for men in question-fields designed for women).
- [Pre-test](#): The data entry programs (the CAPI and the PAPI data entry) are pre-tested by HIS team members; they are then set on Statbel network for a try-out period. Errors or inconsistencies found in the programs are notified to the programmer at Sciensano by e-mail and are corrected.

4.7.4. Quality of data management

See point 4.6. for details.

- [Vertical control](#): Once the data are entered in the database, a series of quality controls are performed. During the vertical control, it is checked whether all information is available for all activated households.
- [Horizontal control](#): Next to the vertical controls, it is necessary to control for internal consistency of the data (through SAS programs developed by HIS researchers per module). Inconsistencies in the data entail going back to the paper-questionnaires for verification. Inconsistencies in data may arise from errors due to the respondent or errors introduced during data entry. Inconsistencies due to the respondents are treated during the analysis (statistical programs for data cleaning). Inconsistencies due to data entry (discrepancy between questionnaires and encoded data) are corrected in the database.
- [Data analysis](#): As stated above, respondent's contradictions or interviewers' mistakes are dealt with in the data cleaning process before the statistical analysis. The programs for both the data cleaning and the analysis are created by the HIS researcher responsible for the given module and systematically verified by another designated HIS researcher.
- [Results](#): The tables of results and the explanative texts produced for a module (= draft version) undergoes internal revision within the HIS team. For some modules, external experts in a given domain assure an external review of the draft report.

4.8. Software development

4.8.1. Description and planning

A member of the HIS team will develop a CAPI (Computer Assisted Personal Interview) application for the face-to-face questionnaire (combination of the household and individual questionnaire). This application serves to interview the respondents and directly encode their answers of the HIS 2023.

This application is developed with the software package Blaise® 5.10.7, a powerful and flexible software. This version is an updated version of Blaise 4.8 which was already successfully used for the development of the CAPI and the data entry programs in the previous HIS cycles.

The Blaise Developer software is installed on the Blaise5serverTest blaise01.corp.darwinproject.be (Remote Desktop Connection). Through its high-speed data-entry and management capacities, Blaise® is suitable for large surveys such as the HIS. This tool is user-friendly as well as for the developer as for the end user (interviewer).

Data-entry can be easily managed by applying specific rules for routing filters (i.e. some HIS questions are not allowed for proxy interviewees, some modules target a specific sub-group of respondents), jumps (depending on the response to a defined question, certain subsequent questions can be skipped), type of questions (open, semi-open, closed questions, multiple response), limiting the number of values, language switching during interview (through menu or with a keystroke), mouse, ... These rules are very useful to limit the number of errors when entering the data during the interviews or the imputation of PAPI (paper self-administered questionnaire) data entry programme (improving the quality of the database).

The CAPI application is developed at Sciensano. After testing, it is integrated in the CAPI application of the contact form at the level of Statbel, on the interviewers' laptop.

The interviewers send the data collected by means of the questionnaires on a regular basis to Statbel server via an online connection. Statbel then exports these data in an adequate format. The database generated in Blaise® has a specific format, but extraction in other formats is possible (ex. TXT). The generation is done in two steps: 1) extraction of the data and 2) extraction of the variables in a SAS format. The advantage of extracting the data in SAS format is that not only the variables and their values are recovered, but also the questions.

Subcontracting (see point 7.3)

Blaise® is also used by Statbel, another reason why we are using this software since we work together with them for the fieldwork. The specific Blaise® 5.10.7 version is used because Statbel also uses this version and so it is already installed on the interviewers' laptop working for other surveys carried out at Statbel. An advantage is that an expertise with this package exists in both institutes.

4.8.2. Validation procedures

- *Vertical control*: Once the data are entered in the database, a series of quality controls are performed. During the vertical control, it is checked whether all information is available for all active households.
- *Horizontal control*: Next to the vertical controls, it is necessary to control for internal consistency of the data (through SAS programs developed by HIS researchers per module). Inconsistencies in the data entail going back to the paper-questionnaires for verification.

Inconsistencies in data may arise from errors due to the respondents or errors introduced during data entry. Inconsistencies due to the respondents are treated during the analysis (statistical programs for data cleaning). Inconsistencies due to data entry (discrepancy between questionnaires and encoded data) are corrected in the database.

- *Data analysis:* As stated above, respondent's contradictions or interviewers' mistakes are dealt with in the data cleaning process before the statistical analysis. The programs for both the data cleaning and the analysis are created by the HIS researcher responsible for the given module and systematically verified by another designated HIS researcher.
- *Results:* The tables of results and the explanative texts products for a module (= draft version) undergoes internal revision. Internal review is done within the HIS team.
- *Handbook for users:*
 - A specific handbook for guiding the interviewer in the use of the PC is made. This handbook contains general information of the use and settings of the PC as well as instructions for using the CAPI program.
 - A handbook for the data-entry programme of the self-administered questionnaire is also available for the operators.

4.9. Website

4.9.1. Sciensano website and link to the HIS page

The easiest way to reach the HIS website is via the direct link:

In English: <http://www.healthsurvey.be>

In French: <http://www.enquetesante.be>

In Dutch: <http://www.gezondheidsenquête.be>

On this HIS website, one can consult some information regarding the survey: funding and collaboration, methods and sampling, protocol and questionnaires, procedure for accessing the Health Interview Survey microdata, reports of the HIS 2018 and 2013, interactive website HISIA, Contact information. During the fieldwork period, one can also find for example application calls for interviewers, specific information for the selected households (folder, video, poster and FAQs), or information from the press (e.g., public announcement of the study).

4.9.2. Web application "HISIA"

External users have the possibility to explore the HIS data in personalised analyses. In this respect, a dynamic tool is made available that allows to compute a wide range of indicators collected through the HIS in function of different background variables. The website tool is called "Health Interview Survey Interactive Analysis" (HISIA): <https://hisia.sciensano.be>

Objective

The objective of developing a flexible web-based application is to facilitate the access and to enhance the use of descriptive statistics generated from the HIS performed in Belgium, for scientific research and public health policy, but also for a broad range of potential users.

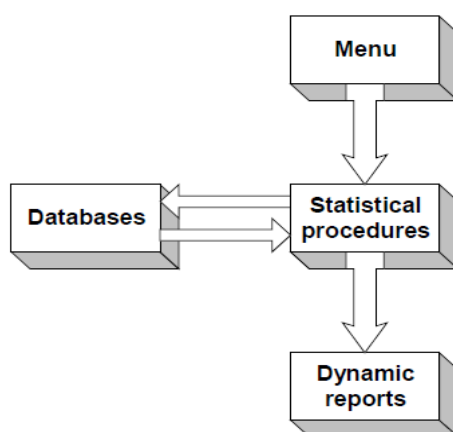
Principle

The principle is to rationalise the statistical programmes for calculating health indicators and run them through parameterisation on the Internet in an interactive way.

General description

SAS® Stored Process was used to create this web-based application. The technical implementation procedures are described in a document stored on the server⁹.

From a practical point of view, to launch the analysis the user first chooses the parameters of interest (i.e., background variables such as the year(s) of survey, region, gender, age group,...) via the *menu system* on the Internet browser. The selected parameters pass from the html-format to the SAS® system, where the corresponding *databases* and *statistical procedures* are invoked via macros. Finally, the browser displays the generated results (*dynamic reports*) in a table format.



4.10. Procedures to obtain the HIS micro data

One of the aims of the HIS project is to deliver a database which can be used for further research. More detailed information about the health survey, the data (use and overview of the variables) and the procedure can be found on the website of the HIS (Procedure for accessing the Health Interview Survey microdata):

<https://www.sciensano.be/en/projects/health-interview-survey#procedure-for-accessing-the-health-interview-survey-microdata>

⁹ Stored on the Sciensano server at: HIS\HIS2023\Research protocol\3 - HISIA.docx

5. Communication of the results and reports

5.1. Reporting mechanism

5.1.1. Sponsor reports

The results of the HIS 2023 (and comparisons with previous survey years) are communicated to the commissioners, the stakeholders and the public via full descriptive reports that represent the five broad domains of interest of the study: Health and quality of life, Lifestyle, Preventive knowledge and practices, Mental health and well-being, Use of health services, Health and society. The results are described at the national level and at the levels of each of the three federated Regions of Belgium, as they have independent health competences and policies, and have co-funded the HIS project. The reports are written, revised and translated (from Dutch to French or vice versa) by the HIS team members. Executive summaries are also produced in two languages, and released to the press for further dissemination.

For ecological and financial reasons, no printouts of the reports are done, but the full electronic reports (in PDF form, per domain and chapter) are available in French (www.enquetesante.be) and Dutch (www.gezondheidsenquête.be) on the HIS website.

Because of the huge workload these reports represent, they are released in a multistage sequence.

5.2. Publication plan: peer-reviewed publications and others

The results of additional analyses are usually presented during scientific conferences, symposiums, national and international meetings, in parliament, or serve to answer parliamentary questions. They are also published in peer-reviewed scientific journals or in medical journals. The list of publications can be found on the HIS website.

5.3. Other forms of communication of results

To disseminate the results of the HIS, a number of communication tools other than the reports are developed in different formats. The format is adapted to the situation and target group. For a wide dissemination of the HIS results among the general public, press releases, press conferences and/or interviews (radio, television) are prepared. Besides, a specific page containing all information about the HIS, including the results, is available on the HIS website (see point 4.10.1). Next to this "static" presentation of reports and summaries, an interactive and user-friendly data analysis tool enables to promote a dynamic analysis of the aggregated data (HISIA, see point 4.10.2) for whom might be interested. Using this instrument, it is possible to produce information on the distribution of any HIS indicator in the population without any statistical skills or programming knowledge. The output is displayed in tables and the indicators can be produced in function of maximum 3 the following background variables: age, gender, household composition, education, income, urbanisation, and province.

In addition to these dissemination tools, a HIS dataset is developed (including a users' manual and codebook) so that the HIS micro-data are available - under specific conditions - for external users (commissioners, academic research teams, foreign research institutions...) to pursue focused analyses. The procedure to obtain the data is depicted on the HIS website (see point 4.11). The database is delivered in any of these four formats: SAS, SPSS, STATA and ASCII. These files include the variables

and indicators constructed using the 2023 data, and when possible, the same variables and indicators from previous HIS surveys.

6. Risk and benefits for participants (Ethics)

The risk of harm to those who take part in a questionnaire-based survey like the HIS is very limited. The greatest risk for the respondents could be caused through the disclosure of their personal data, so special attention is devoted to the protection of the information collected, both on the short term (encrypted data on interviewers' laptop, forbidden disclosure, anonymisation, ...) and on the long term (secure data transfers and storage, limited access, recoded data,...).

The benefits from participating in the HIS are not directly discernible for the respondents, but are to be conceived at the community level, in the perspective of developing public health policies that fit the needs of the given population, based on the evidence of research.

The basic rights of the (potential) participants are generally guaranteed through different "safety" practices and gatekeeping procedures in health and epidemiologic research. The following measures are ensured in conducting the HIS:

Ethical committee

The HIS 2023 research protocol is submitted to the Ethical Committee of the University of Ghent. The survey takes place under the condition that Sciensano is granted permission to proceed. The Committee gives an approval if all aspects of the study are ethically sound and meet the best practice guidelines.

Information to the participants and consent

In previous HIS waves, no formal written informed consent was used as it was deemed unnecessary because the research is carried out in settings that pose no threat, the participation is voluntary and it is obvious that no benefits are lost if invitees refuse to take part. This situation arises in "naturalistic" studies such as the HIS, based on questionnaires or interviews where providing the requested information involves giving *de facto* consent. Still, for the HIS 2023, it was decided to ask for a written informed consent.

There are 3 informed consent forms: One for the adults aged 18 years and older, one for the parents and guardians of children and one for children aged 12 to 17 years old. In case of a proxy interview (for an adult person incapable of responding for him/herself), the proxy respondent will be asked to sign the adult informed consent form on the behalf of the participant.

As was the case for previous cycles, providing complete information to the invitees as to the expectations regarding the survey is a deontological principle honoured in the HIS. The HIS ensures the following practices:

- Households that are selected to participate in the HIS receive a personalised invitation letter as well as a notification leaflet about the survey itself, the aims, the practical organisation (with a link to a demo video), the institution in charge, the commissioners of the survey, when and how the interviewer will contact them, the time to fill out the questionnaires, and an example of results from previous surveys.
- It is clearly stipulated that participation is voluntary and that withdrawal is possible at any time. Anonymity of the disclosed information is guaranteed.

- An e-mail address, Internet website and free telephone number are clearly indicated on the leaflet if selected households need further information or don't want to be contacted by the interviewer. Potential participants can also ask more about the survey or refuse to participate at the moment the interviewer contacts them to ask their consent and make an appointment.
- At the time of the interview, participants are told that they can refuse to answer specific questions or that they can end the survey without prejudice or a need for justification.

Confidentiality, anonymity and data security

Measures are taken in the HIS to certify that participants cannot be identified and that their responses remain confidential. These measures are described in point 4.7 "data security measures" of the present document. Other protection measures include the following:

- Interviewers sign a confidentiality clause inserted in their contract. Interviewers are forbidden to disclose any information gathered during the interview or make any general inference on the basis of the collected information.
- The employees responsible for encoding the HIS data from the self-administered questionnaires at Statbel are submitted by Decree (M.B. 20.07.62 §2) to an obligation of professional secrecy concerning personal information.
- Data are coded: Sciensano researchers do not have access to the name, address or national register number of selected households, which are kept at the secretariat of Statbel for fieldwork management. Only coded data are transferred to Sciensano researchers. Small cell analyses have been performed on data of previous HIS and show minor risk of identification.
- Sciensano researchers have an obligation of confidentiality by contract or by status. Besides, only the researchers implied in the HIS team have access to the data (protected network emplacement).
- External researchers, who purchase the anonymised/restricted/aggregated HIS database to carry out scientific analysis, sign a bilateral contract where a confidentiality clause is stipulated as well as the obligation to renounce to identify individuals or transfer the dataset to a third party.

Communication of results

All results derived from the HIS are divulged in a format that impedes any recognition of individuals, i.e. in tables of aggregated statistics, graphs (See previous reports on website: www.healthsurvey.be).

What in case of violation of data security measures

If interviewers disclosed any given information that is used or collected during the HIS fieldwork, this would constitute an infringement to the Law of July 4, 1962 (Art. 18 – 22) relative to public statistics and would be penalised according to the disciplinary proceedings established in the article 458 of the Penal Code. They would also be fired from the position of HIS interviewer.

If unauthorised individuals force the access to the HIS data, there is no chance these could be used to harm participants. Indeed, the data are kept in an anonymised format and the logical key is not available at Sciensano. Any attempt to recover the national number of the participants would be vain, as the third trusted party holds the logical key for possible data linkage, but is never authorised to give

it away. Small cell analyses also revealed that the identification of respondents on the basis of the existing data is practically unfeasible.

7. Organisation of the research project

7.1. Starting and completion date

The project consists of three different phases, roughly described hereunder, starting 01/01/2022 and ending 31/12/2024:

01/01/2022 – 31/12/2022	Preparatory phase: conceptual papers, questionnaire development, pre-test and adjustment, programming and printing questionnaires, invitation letters, information leaflets, training of - and manuals for - interviewers, ethical and privacy application documents, etc.
01/01/2023 – 31/12/2023	Fieldwork: data collection phase, field monitoring, data management, consistency check programs, supply of information to interviewers...
01/01/2024 - 31/12/2024	Data encoding, data cleaning and management, data analysis and reporting, dissemination of results, web program for interactive analysis, database for external users. This phase can be prolonged beyond 2024.

7.2. Subcontracting Statbel

Activities related to the HIS 2023 data collection are subcontracted to Statbel. In this context, a contract has been established and signed – after approval by the juridical department and DPO of Sciensano – between the general directors of Sciensano and of Statbel. In this contract, the tasks and responsibilities of Statbel regarding the HIS data collection are specified.

The on-going ‘state of the art’ regarding the fieldwork progression on the basis the results obtained from the “data collection control procedures” (or the paradata) are discussed during weekly meetings between the delegates from Statbel and Sciensano. In case a problem occurs, a solution is sought in mutual agreement. Such meetings assure that Sciensano has a clear representation of the data collection progression and is able to verify if all Statbel subcontracted activities are being carried out according to the contract.

7.4. Scientific review

The HIS team shares experience with an advisory scientific board and ad hoc consultations with expert in academia, health policy spheres or from the field.

Scientific Steering Committee of the project

A Scientific Steering Committee (SSC) is set up to assure the scientific follow-up of the HIS 2023. The role and tasks of the SSC are detailed in the Agreement Protocol signed by the Commissioners of the HIS. Its main task is to advice the Commission of Commissioners on scientific issues related to the survey. Meetings with the SCC are organised at least once a year during the three years of the project. The HIS team members expose the current state of affairs and potential difficulties or points of discussion for the proper development of the project. The minutes of these meetings are sent to all Commissioners for information, or further processing if necessary.

Expert consultations

In defining the content of the HIS 2023 questionnaires, experts and members of health agencies are consulted to discuss the content of modules that require revision. It is to mention that most modules will be kept unchanged from the previous surveys to ensure trend analysis, some modules need an in-depth revision, and some new modules are added to the questionnaires.

On-demand scientific review

The HIS is a complex survey with various scientific disciplines involved. Despite a detailed planning, unexpected problems can emerge throughout its execution, as well as during the analysis of data where new research orientations or unpredicted findings can require the input of the scientific community. Thanks to the many formal and informal contacts the HIS-team has built since the start of the survey, scientific advice can easily be solicited.

7.5. Propriety rights of study material and results

All data collected by means of the HIS 2023 are the property of the Commission of Commissioners, as mentioned in the Inter-ministerial Agreement Protocol. The results that stem from the analysis of the HIS data produced at Sciensano are first communicated to the Commissioners (owners of the data) by means of reports. The Commissioners receive the database at the end of the reporting process.

7.6. Client satisfaction

7.6.1. Definition of the Clients

The clients of the HIS are:

The Commission of Commissioners

The Commission of Commissioners is the sponsor of the HIS. The tasks of the Commission are described in the Agreement Protocol. All aspects of the survey (methodology, content, analysis, report) are discussed and approved by the Commission. When a consensus is reached on the different points of discussion, the conclusions are transferred to the so-called Policy Coordination Working Groups (formerly 'Inter-cabinet Working Groups') and communicated to the Inter-ministerial Conference Public Health.

The sponsors of an oversampling

Provinces or Communities can sponsor an oversampling within their geographical entity, but they will not be represented as such in the Commission of Commissioners. For the HIS 2023, an oversampling is planned in the German Community areas (East Belgium).

The users of the HIS data

Once the Commissioners have received the results of all topics of the HIS 2023, the micro-data are prepared, together with an users' manual and a codebook, so that external researchers can exploit the data in accordance to their own domain of interest. The data transfer procedures are described on the HIS website.

7.6.2. Contacts with the Clients

The Commission of Commissioners

Contacts with the Commission are made through meetings. There is no set schedule for these meetings, but there should be a minimum of one per year. In-between meetings, the members of the Commission will receive via e-mail the necessary information allowing them to have an updated insight of all aspects related to the survey. If necessary, any member of the Commission can ask for supplementary information.

The sponsors of an oversampling

Every trimester, contacts with the German Community of East Belgium are foreseen via e-mail in order to inform them on the number of interviews completed in the area. In case of potential difficulties, meetings can be organised to find relevant solutions.

The users of the HIS data

The Commissioners and the members of the Scientific Steering Committee, but also academics and researchers that have been using the HIS data in the past, are contacted to let them know when the HIS 2023 micro data are ready for dissemination. The process to access the data is described on the HIS website and all related documents are available online.

7.6.3. Satisfaction inquiry of the client

First, commissioners and sponsors are in contact with the HIS team members during meetings, but also, through e-mails for specific requests. The external researchers and academics are represented in the Scientific Steering Committee, where recommendations are given to meet the expectancies of the community. Feedback and satisfaction of requirements is thus an on-going process throughout the conduct of the project.

7.6.4. Treatment of complaints

In accordance with: SOP I/03/22/NF.

7.7. Archiving process

All the HIS documents and data files are kept on the FS Volume of the secured internal network server. At the start of each new HIS project a new subdirectory is created. Each HIS is classified with reference to the year it is accomplished. More specifically, the HIS 2023 subdirectory is placed in: \\sciensano.be\FS\1140_DATA\HIS\HIS2023

8. References

- (1) Research strategies for health based on the technical discussions at the 43th World Health Assembly on the role of health research in the strategy for Health for All by the year 2000. New York: Lewiston; 2008.
- (2) Van Oyen H, Tafforeau J. Health Interview Survey. Arch Public Health 1994; 52:79-82.
- (3) Van Oyen H, Tafforeau J, Hermans H, Quataert P, Schiettecatte E, Lebrun L, et al. The Belgian Health Interview Survey. Arch Public Health 1997; 55:1-13.
- (4) Bergner M. Measurement of Health Status. Medical care 1985; 23(5):796-804.
- (5) De Kleijn-De Vrankrijker MW. Internationale aspecten van gezondheidsmeting. In: Gunning-Schepers LJ, Mootz M, editors. Gezondheidsmeting. Houten: Van Loghum; 1992.
- (6) Mootz M, Van den Berg J. [Indicators of health status in the CBS-Health Interview Survey]. Maandbericht Gezondheidsstatistiek CBS 1989; 2:4-10.
- (7) Guide to Clinical Preventive Services: Report of the US Preventive Services Task Force. 2 ed. Baltimore: William & Wilkins; 1996.
- (8) Schaapveld K, Bergsma EW, Van Ginneken JKS, Van De Water HPA. Setting priorities in prevention. Leiden: TNO Institute for Preventive Health Care; 1990.
- (9) Winslow CEA. The untilled fields of public health. Science 1920 Jan 9; 51(1306):23-33.