

BELCOHORT

Opportunities for a population-based cohort in Belgium

NIENKE SCHUTTE • JOHAN VAN DER HEYDEN • BRECHT DEVLEESSCHAUWER •
KARIN DE RIDDER • HERMAN VAN OYEN

WHO WE ARE

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

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

Sciensano

Epidemiology and public health - Lifestyle and chronic diseases

Belcohort

Sept 2020 • Brussels • Belgium

Internal reference number: D/2020/14.440/84

—

Nienke Schutte

•

Johan Van Der Heyden

•

Brecht Devleesschauwer

•

Karin De Ridder

•

Herman Van Oyen

•

Please cite as: Nienke Schutte, Johan Van Der Heyden, Brecht Devleesschauwer, Karin De Ridder, Herman Van Oyen. BELCOHORT: Opportunities for a population-based cohort in Belgium. Brussels, Belgium : Sciensano ; 2020. Report number: D/2020/14.440/84.

EXECUTIVE SUMMARY

Ever since Frost coined the term 'cohorts' (in 1935), referring to his study on age specific mortality rates and tuberculosis, this research design was embraced by many epidemiologists to capture exposure-outcome relations in specific populations. The determinants of a healthy lifestyle as well as healthy ageing are all interrelated and result from complex interactions between a wide range of environmental exposures, phenotypic characteristics and genomic factors. Population-based cohort studies are key in providing longitudinal data on (the impact of) major determinants of health and disease and provide a sound base for targeted policy, policy follow-up and evaluation of the health and social systems. Recently, the COVID-19 crisis showed that the rapid exchange of, and access, to up-to-date health data is important for rapid response to research questions that concern current public health challenges. In the aftermath of the crisis, information is needed on social, economic, cultural and physical environment (including behavioral data, food, education, work environment, sickness leave/return to work, housing, school) in order to study the effects on health status and population health.

In December 2016, the idea of a cohort as research infrastructure in Belgium has been discussed during a meeting with different Belgian stakeholders (researchers, BELSPO, RIZIV/INAMI, StatBel, KCE, IMA/AIM, regional authorities), agreeing on the importance of investigating the reflection on the relevance, feasibility, cost effectiveness and expected outcomes. In 2018 the Health Working Group for the Strategic Investment Pact (established by then Prime Minister Charles Michel) stated in their report that 'in line with developments in other countries, Belgium will need a cohort of its population as a research infrastructure'. Subsequently, a project aiming to map the current knowledge gaps, include stakeholders and explore the prospects for establishing a population-based cohort was launched in late 2018.

The findings of this study are describes as opportunities for Belgium create a unique take on adding a cohort-live initiative to the public health landscape. Many opportunities to set up a population-based cohorts exist, albeit perhaps not in the classical sense of a cohort, such as the LifeLines in the Netherlands or CONSTANCES in France. These opportunities should be explored in depth, consolidating collaboration between different stakeholders and sustainability of the resulting research infrastructure. Belgium, being a country with many authorities, services and agencies, but short communication lines, might hold the right cards to build an infrastructure holding longitudinal data collections of register data, survey data and/or clinical data.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	4
ABBREVIATIONS	7
BACKGROUND	9
1. Context.....	9
2. Cohort initiatives in Europe	9
3. Success factors of a population-based cohort	13
4. In Belgium.....	14
5. Opportunities.....	15
RELEVANCE OF A POPULATION-BASED COHORT STUDY IN BELGIUM	17
1. Contributing factors.....	17
1.1. Health across the lifespan.....	17
1.2. Socio-economic factors.....	18
1.3. Upbringing and household environment	19
1.4. 'Omics'	19
2. Identifying the correlations	19
3. Availability longitudinal (health) data.....	21
4. Belgium-specific (health) policies and interventions.....	21
5. Position in the European landscape	22
6. The case of COVID	23
OPPORTUNITIES IN BELGIUM	25
1. National Health Data Platform	25
1.1. Background.....	25
1.2. Concept	28
1.3. Impact & Risks.....	30
2. Online panel.....	32
2.1. background	32
2.2. Concept	33
2.3. Impact & Risks.....	35
3. Other Opportunities	37
3.1. Multiple initiatives, one portal	37
3.2. Technological advances in large scale longitudinal research	38
RECOMMENDATIONS	41
CONCLUSIONS	43
REFERENCES	45
ANNEXES	49
1. Key meetings.....	49
1.1. Stakeholders.....	49
1.2. Summary key meetings	50
2. Workshop.....	55
2.1. Agenda	55
2.2. Sessions	55
2.3. Conclusions (Summary).....	56

ABBREVIATIONS

EPS	Échantillon Permanent(e) Steekproef
ESS	European Social Survey
FWO	Fonds voor Wetenschappelijk Onderzoek
HES	Health examination study
HIS	Health interview survey
IMA/AIM	L'Agence Intermutualiste - Het InterMutualistisch Agentschap
ISSEP	Institut Scientifique de Service Public
KCE	Federaal Kenniscentrum - Centre fédéral d'expertise
RIZIV/INAMI	Rijksinstituut voor ziekte- en invaliditeitsverzekering/Institut national d'assurance maladie-invalidité
SHARE	The Survey of Health, Ageing and Retirement in Europe
StatBel	Belgium Statistics
VITO	Vlaamse Instelling voor Technologisch Onderzoek
WHO	World Health Organisation

BACKGROUND



1. Context

As life expectancy increases in the western countries, new challenges in the field of public health present themselves. The number of patients with cancer, stroke, chronic respiratory disease, and neurologic or psychiatric disorders such as dementia and depression are increasing dramatically due to the rapid lengthening of life expectancy, together with limitations in daily life. Whereas the quality of the clinical care for these conditions has improved enormously, the increase in healthy life years is lagged behind, but not homogenous throughout society and divergent evolution are observed by gender and/or social position (Renard et al., 2019; Van Oyen et al., 2011). Healthy ageing does not only mean being free of age-related chronic disease, but also includes physical and cognitive capability and general feelings of well-being. At the same time among the younger generations, unhealthy food consumption behavior, the lack of physical activity and other lifestyle factors such as smoking and alcohol consumption are leading to an increasing prevalence of obesity and a growing number of patients with diabetes, cancer and cardiovascular diseases. These challenges make their mark on the political agenda, addressing for example the resilience of the health care and social services or the viability of the pension systems.

The determinants of a healthy lifestyle as well as healthy ageing are all interrelated and result from complex interactions between a wide range of environmental exposures, phenotypic characteristics and genomic factors across the lifespan. Large, population-based cohort studies, in which individuals are followed for a defined period in which a variety of variables can be measured, are vital to ensure sufficient statistical power to better understand the role of various environmental factors (social, physical, etc.) and their interaction with complex traits. These cohorts provide great opportunities to assess longitudinal exposure-outcome relations on a large-scale and provide insights in the path to disease or other health outcome, allowing to evaluate markers for early detection of disease, to identify individuals at increased risk for developing (chronic) diseases, to unravel the complex interaction between environment and genetics and to assess drivers for socioeconomic disparities in health. The evaluation of multiple hypotheses and its external validity (i.e. the applicability of its results to a defined population) justifies the high costs and logistic complexities of a population-based cohort.

2. Cohort initiatives in Europe

Since the 1980s, large population-based cohorts have been established all over Europe. Among the largest cohorts, health and socioeconomic factors are the main focus. Table 1 provides an overview of the most important multipurpose population-based cohorts in Europe with sample sizes of more than 50.000 individuals. These cohorts differ in origin. The set-up or establishment of a cohort depends on several factors. From a practical point of view, the data that is already available may provide an obvious

starting point. For example, the Norwegian CONOR study combined all core survey data and stored blood samples from 10 large regional cohorts (see also Box - The birth of a cohort) into one population-based cohort. The Swiss National Cohort is based on census data from 1990 and 2000 that were linked to mortality, life birth and emigration records, thereby comprising the whole population in the country. These cohorts provide excellent examples of combining available data, thereby establishing (passive) prospective study cohorts. In France, everyone with health insurance is entitled to receive free health examinations that include extensive work-ups (see Box). Researchers, armed with the knowledge gained from the GAZEL Cohort Study (established in the late 1980s, comprising employees of a utility firm in France), used this existing infrastructure to set-up an epidemiological research infrastructure. Eligible persons were randomly selected to come to their Health Screening Center for health examinations and to fill out questionnaires. These data can be linked to health and socioeconomic national databases, thereby creating a population-based multipurpose cohort.

In other countries, administrative data is not centralized nationally, but rather managed by the different regions or provinces. In these countries, starting from an existing (data-)infrastructure is not always possible. For example, inspired by the successes of the European Prospective Investigation into Cancer and Nutrition (EPIC) Cohort in 1992, researchers in Germany called for a similar project to be set up. However, the availability of administrative data on a national level was scarce: data is mainly collected on state-level (*Länder*) or in private companies (for example private health insurance companies). Rather, a close collaboration between universities (funding by the states), the Helmholtz Institutes (federally funded) and the federal German government resulted in an extensive data collection in 200.000 individuals.

In addition to data that is already available or infrastructures that that be used or employed to set up a cohort, a specific aim or research question also stipulates the set-up of a cohort. Many cohorts focus on health or healthy ageing, life style factors and the path to disease (HUNT, CONOR, UK-Biobank, LifeLines, LifeGene, CONSTANCES, GNC). Every cohort, however, also aims to formulate a unique selling point. For instance, in Estonia, the focus lies on health in relation to genetic research. In their biobank, genetic information of over 52.000 individuals is stored.

Largely, European population-based cohorts employ three ways to recruit participants. As population-based cohorts aim to be representative of the targeted population, three cohorts used sampling frames created by the national registries i.e. HUNT, LifeGene and GNC. In an effort to increase response rate and reduce loss to follow-up, two cohorts recruited participants by means of the GPs; the Estonian Biobank and LifeLines. In the latter cohort, 49% of the included participants were invited through their GP, whereas 13% self-registered via the LifeLines website. The remaining 38% comprised of family members of the invited individuals. The participants in the LifeGene cohort and CONSTANCES are also encouraged to invite their family members and/or members of their household. The response rate of the cohorts shows large differences, ranging from 88% for the first wave of the HUNT cohort (HUNT 1) to only 5.5% for the UK Biobank. For the HUNT study, the high response rate might be attributed to the focus on one geographical area of Norway (24 municipalities in the Nord-Trøndelag County), which can

aid in the word-of-mouth. For the more recent cohorts (LifeGene, CONSTANCES and GNC), calculating a response rate was not possible as the first wave of data collection is still ongoing. All cohorts in Table 1, except for the Swiss National Cohort, which includes the whole population, show overall good representativeness of the general population. For many of the cohorts (e.g. Estonian Biobank, UK Biobank and LifeLines), participants were more likely to be female, less likely to be obese and to smoke on a daily basis and have higher socioeconomic status than nonparticipants, suggesting the existence of a healthy volunteer bias.

Largely, European population-based cohorts employed three ways to recruit participants. As population-based cohorts aim to be representative of the targeted population, three cohorts used sampling frames created by the national registries i.e. HUNT, LifeGene and GNC. In an effort to increase response rate and reduce loss to follow-up, two cohorts recruited participants by means of the GPs; the Estonian Biobank and LifeLines. In the latter cohort, 49% of the included participants were invited through their GP, whereas 13% self-registered via the LifeLines website. The remaining 38% comprised of family members of the invited individuals. The participants in the LifeGene cohort and CONSTANCES are also encouraged to invite their family members and/or members of their household. The response rate of the cohorts shows large differences, ranging from 88% for the first wave of the HUNT cohort (HUNT 1) to only 5.5% for the UK Biobank. For the HUNT study, the high response rate might be attributed to the focus on one geographical area of Norway (24 municipalities in the Nord-Trøndelag County), which can aid in the word-of-mouth. All cohorts in Table 1, except for the Swiss National Cohort, which includes the whole population, show overall good representativeness of the general population. For many of the cohorts (e.g. Estonian Biobank, UK Biobank and LifeLines), participants were more likely to be female, less likely to be obese and to smoke on a daily basis and have higher socioeconomic status than nonparticipants, suggesting the existence of a healthy volunteer bias.

Table 1. Overview and sample size of ongoing European multipurpose population-based cohorts with a sample size >50.000.

Name	Country	Year	Sample size	Age of the sample ^a	Aim	Reference
HUNT	Norway	1984	126.000	13-100yrs	Health related lifestyle, prevalence and incidence of somatic and mental illness and disease, health determinants, and associations between disease phenotypes and genotypes.	(Krokstad et al., 2013)
CONOR	Norway	2002	174.430	20-103yrs	To describe Norwegian men and women in terms of distribution of exposures and health status according to time, place and socio-economic factors.	(Naess et al., 2008)
Estonian Biobank	Estonia	2002	52.000	18+yrs	Collect information on the health status of the Estonian population combined with genetic information and to use the results of genetic research to improve public health.	(Leitsalu et al., 2015)
Swiss National Cohort	Switzerland	2005	^b 7.452.075	All ages	To build up, maintain and yearly extend the first nationwide cohort; provide an evolving, multi-faceted research platform and offer other cohorts and clinical databases expertise in record linkage and information on socio-economic status.	(Bopp et al., 2009)
UK-Biobank	UK	2006	500.000	40-69yrs	To investigate why some people develop particular diseases while others do not; to understand the causes of diseases better, and to find new ways to prevent and treat many different conditions.	(Sudlow et al., 2015)
LifeLines	The Netherlands	2006	167.729	0-93yrs	To establish a resource for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy ageing.	(Scholtens et al., 2015)
LifeGene	Sweden	2010	^c 200.000	0-50yrs	To build up a resource for research in all medical disciplines, enabling new and groundbreaking research on the relationships among heredity, environment and lifestyle.	(Almqvist et al., 2011)
CONSTANCES	France	2012	^c 200.000	18-69yrs	To serve as an epidemiological research infrastructure accessible to the epidemiologic research community with a focus on occupational and social factors, and on chronic diseases and aging.	(CONSTANCES team et al., 2015)
German National Cohort (NAKO)	Germany	2014	^c 200.000	20-69yrs	To investigate the causes for the development of major chronic diseases, i.e. cardiovascular diseases, cancer, diabetes, neurodegenerative/-psychiatric diseases, musculoskeletal diseases, respiratory and infectious diseases, and their pre-clinical stages or functional health impairments.	(German National Cohort (GNC) Consortium, 2014)

Note. ^aat inclusion; ^bresidents in Switzerland who participated in the 2000 population census; ^cEstimations, as recruitment for the first wave is ongoing.

The birth of a cohort: 3 examples

CONSTANCES

In France, everyone with health insurance from CNAMTS (“Caisse Nationale d’Assurance Maladie des travailleurs Salaries”), as well as their dependents, is entitled to receive free health examinations that include extensive work-ups conducted in Health Screening Centers (HSCs). Scientists from INSERM, experienced in short studies such as GAZEL cohort study, interested in occupational and social factors, on chronic diseases and ageing, initiated a partnership with CNAMTS, in 22 selected HSCs located in 19 ‘departements’ in different regions of France, eligible persons were randomly selected to come to their HSC for examination. By 2019, 200.000 individuals are included in the CONSTANCES cohort.

LifeLines

In the early 2000s, the Dutch government cancelled the plans for building a high speed railway connection connecting the northern part of the Netherlands with the big cities in the western part of the country (‘Randstad’). The Board of Directors of the University of Groningen, together with the University Medical Center Groningen (UMCG) secured this budget for setting up a population-based cohort in the northern part of the Netherlands (provinces of Groningen, Friesland and Drenthe). Additional funding was provided by the University of Groningen, UMCG and the provinces and in 2006, nearly 100m was secured for 10 years of LifeLines cohort, including two waves of extensive data collection.

CONOR

In Norway, a number of large population-based cardiovascular surveys had been conducted since the beginning of the 1970s. As all surveys comprised a common set of questions, standardized anthropometric and blood pressure measurements and non-fasting blood samples, a new cohort was established including both core survey data and stored blood samples. In 1994 the Norwegian Ministry of Health appointed the Steering Committee for the CONOR collaboration.

3. Success factors of a population-based cohort

To ensure the sustainability of a cohort, several success factors can be identified based on the successes (and failures) of other cohorts. One of the most important factors is the *impact* (of the results) of the population-based cohort. A cohort on a larger scale requires a great deal of commitment and awareness raising and should address the societal challenges in the field of public health which can concern demographic change, wellbeing, environment, freedom, security etc. In addition, impact of cohort can be evaluated by its significance for scientific research and policy makers. The data collected should be essential for providing answers to the burning scientific questions as well as policy reports and evaluation of health strategies. The *position of the cohort in the health information system of a country* characterizes the importance of the link between data collected within a cohort and the national strategies. A cohort should fit in the existing research infrastructures, but at the same time possesses unique features.

However, in order for data to be used for these purposes, data should be easily *accessible* for these parties in terms of low costs and speed at which data is made available. This requires monitoring and documentation of the quality and reliability of the data. This calls for a clear governance structure and data management plan. All of the large population-based cohorts in Europe have policies to share their data. Generally, a research proposal should be prepared and submitted to the Scientific Committee of

the cohort for review. For most cohorts, only observational study designs are eligible. The call for research projects excludes the interventional studies (e.g. CONSTANCES, LifeLines, HUNT). In some cases, nested observational studies are welcomed, but results should be made public through publication in scientific journal or reports with a statement that the research was conducted in that specific cohort. Some cohorts encourage or even require that the name of the cohort appears in the title of every publication (e.g. CONSTANCES, UK Biobank). Some cohorts require research groups from abroad to collaborate with a national researcher to apply for data (e.g. HUNT).

Furthermore, the *availability of sustainable funding* can be listed as a factor for success. The realization of a cohort requires a large financial and organizational effort from the start. Yet, generally the funding issues of large cohorts emerge some years after the establishment of the cohort. Stable long-term funding increases the value of longitudinal data exponentially with the number of study waves. In Europe, many countries split up funding for scientific research in smaller grants, which are awarded in competitive application processes. This results in an inability to allow for long-term research goals and structural data management. Consequently, interdisciplinary biomedical-sociological longitudinal studies experienced difficulties in securing funding in the long run, despite having built an excellent infrastructure for epidemiological research. The largest European population-based cohorts list their federal government (and/or specific ministries) as their main source of funding, together with national research councils and academic institutes. UK Biobank also lists charity organization such as Diabetes UK and the British Heart Foundation. Although the HUNT has largely been publicly funded, funding is also been provided by commercial enterprises, such as Statoil, and GlaxoSmithKline.

Finally, a success factor of such an initiative, is *excellence*. Amongst others, this can be defined by the establishment of a competent interdisciplinary staff. A high level of scientific qualification amongst the cohort's academic management and the continuous training and motivation of the employees involved are key conditions of success. Furthermore, as cohorts involve topics such as biology, sociology, health, as well as statistical methods of data analysis (including the analysis of longitudinal data and multi-level analyses), knowledge of meta-analytical techniques and in the field of large-scale data processing, the availability of cohort data can support the education programs in higher education and might serve as a breeding ground for new talents. Finally, excellence can be pursued by collaboration and networking with other research groups, (academic) institutes within the country and abroad as well as (local) authorities and administrators and health organizations is eminent for the survival. The involvement of citizen scientists and crowdsourcing is also important to take into account. The success of a cohort will depend on the exploitation of these issues, as this ensures a large impact.

4. In Belgium

Belgium has some cohort-like initiatives. The most extensive cohorts are part of larger European initiatives. A good example is the Survey of Health, Ageing and Retirement in Europe (SHARE); a multidisciplinary and cross-national panel study, which is conducted biannually since 2004, currently including 7 complete waves of data collection (Börsch-Supan, 2019). By collecting data on health, socioeconomic status, and social and family networks from individuals aged 50 and older and their

partners, it strongly contributes to the understanding of the ageing process in the 20 participating countries. All data are collected by face-to-face, computer-aided personal interviews (CAPI), supplemented by a self-completion paper and pencil questionnaire. All SHARE respondents who were interviewed in any previous wave (including nonresponding partners) are part of the longitudinal sample, ensuring a longitudinal design (refreshment samples are drawn regularly to compensate for the reduction in panel sample size due to attrition and maintain representation of the younger age-cohorts of the target population). Another European initiative is the Statistics on Income and Living Conditions (SILC), which contains both a cross-sectional panel, with variables on income, poverty, social exclusion and living conditions, and a longitudinal data panel pertaining to individual-level changes over time. Other European studies such as the European Social Survey (ESS) exploit cross-sectional designs.

On national level, there is some experience in population-based cohorts. For example, the Interface Demography Research Group of VUB does research activities based on a mortality follow up of the population census 1991, 2001 and 2011. The database used by the Interface Demography is managed by Statistics Belgium. Some smaller initiatives focus on specific subgroups of the population, such as patients (e.g. HIV cohort, Van Beckhoven et al., 2012), twins (East Flanders Twin Survey (Derom et al., 2002)), children (Limburgs geboorte cohort, (Janssen et al., 2017)). Finally, IMA/AIM governs a large population-based sample, the Echantillon Permanente Steekproef (EPS), which contains a randomly sampled cohort of health insured individuals in Belgium. The database consists of three types of data, one regarding the population, one containing a database with reimbursed health care procedures and one with the reimbursed medication.

However, some of these have important limitations. Often there is no active recruitment and follow-up is only passive. This is mainly the case for cohorts that are based on administrative databases. Furthermore, some cohorts focus on specific population groups such as patients, specific occupational groups or specific age groups. The scope of each of these studies is limited, based on a well-defined specific research questions with a specific exposure and a limited number of outcome measures. Cohorts that exist for example within an academic environment, often linked to individual research groups with a closed user's network and data might not always be available for other parties. Finally, for some studies a systematic link of the study results with health policy is missing.

5. Opportunities

In December 2016, the idea of a cohort as research infrastructure in Belgium has been discussed during a meeting with different Belgian stakeholders (researchers from almost all Belgian universities belonging to several departments, BELSPO, FWO, RIZIV/INAMI, StatBel, KCE, IMA/AIM, regional authorities). There appeared to be a lot of interest in this initiative and the idea to start up a reflection on the relevance, feasibility, cost effectiveness and expected outcomes. In 2018 the Health Working Group for the Strategic Investment Pact (established by Prime Minister Charles Michel, aiming to provide solid advice on urgent investments in Belgium) stated in their report that 'in line with developments in other countries, Belgium will need a cohort of its population as a research infrastructure'.

Finally, a project aiming to map the current knowledge gaps, include stakeholders and explore the opportunities for establishing a population-based cohort was launched in late 2018. In 2019, several stakeholder meetings were organized together with an international work shop on cohorts studies, with stakeholders representing policy makers at federal level, researchers from most Belgian universities, RIZIV/INAMI, StatBel, KCE and IMA/AIM and many other interested parties. Table 2 shows a brief timeline of the project. Annex I lists the main conclusions of the stakeholder meetings, whereas Annex II provides a report of the international work shop on cohort studies.

Table 2. Brief timeline of the project.

December 2018	First general meetings with stakeholders
December 2019 – March 2019	Brainstorm sessions
January 2019 – March 2019	(virtual) interviews cohorts abroad
March 209 – June 2019	In depth discussions work groups stakeholders
September 2019	International Workshop Cohort Studies
January 2020 – May 2020	Consolidation results, set-up spin-off projects

This report summarizes the main conclusions of these meetings and the workshop and provides an overview of the opportunities within the Belgian landscape for a population based cohort.

RELEVANCE OF A POPULATION-BASED COHORT STUDY IN BELGIUM

1. Contributing factors

The determinants of health across the life span (lifestyle, medical interventions and ageing) are all interrelated and result from complex interactions between a wide range of environmental exposures, phenotypic characteristics and genomic factors across the lifespan. Population-based cohort studies constitute an essential element of all research, aiming to identify the conditions of a healthy development and successful ageing, and social environment and a favorable economic situation over the life course. Figure 1 depicts the general contributing factors and stresses the need for the identification of correlations between these factors.

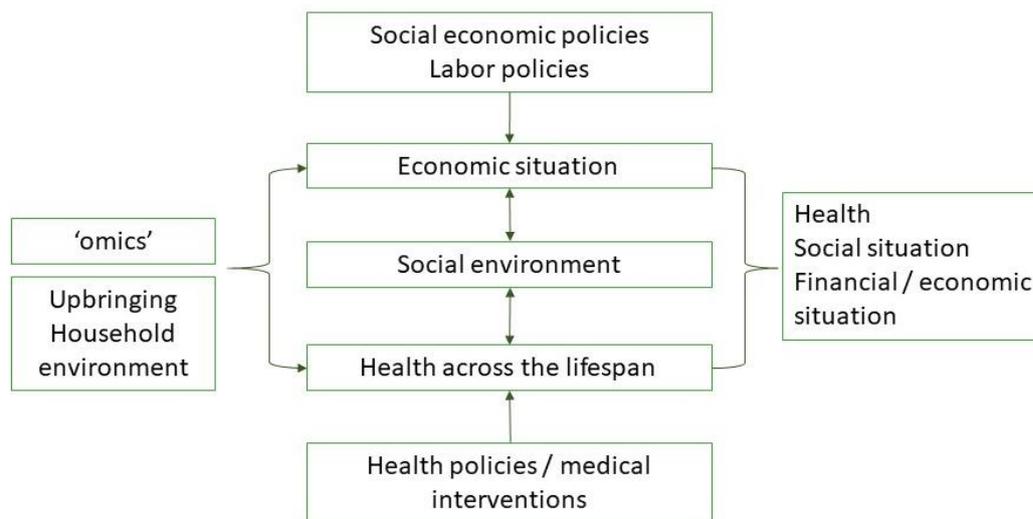


Figure 1. Factors and their interactions contributing to a healthy development and social environment and a favorable economic situation

1.1. HEALTH ACROSS THE LIFESPAN

According to the WHO, *health* is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity. Following this broad definition makes measuring health in general a challenge. Fortunately, derivatives of physical wellbeing or the absence of diseases can be measured, such as hospitalizations, medication, population screenings (to detect for example cancers), etc. and in theory, this data can be analyzed in a longitudinal design, to track health across the lifespan. In Belgium, the existence of a national register number might allow for tracking of data from health registries, such as reimbursements or hospitalizations. However, in practice this data is scattered over different regions and institutions. Moreover, not all health data is organized in registries (e.g. GP data) and few information is available from individuals that are healthy and rarely seek medical assistance or

care. Finally, administrative sources cannot be used to collect data on lifestyle. However, they are direly needed to determine which actions or interventions are necessary to *prevent* individuals from falling ill, becoming disabled or distressed.

The definition of WHO also includes other components, such as mental and social well-being. Health care, which is in the classical sense a field that is dominated by preventing and curing diseases, could profit by integrating the principles of well-being and salutogenesis into their practice and management. Especially when the population is ageing, a salutogenic thinking might make individuals not only live longer, but enjoy a better quality of life, and, as a consequence, are less of a burden to society.

Health surveys are valuable components of a health information system as they provide simultaneously information on physical, mental and social health, lifestyle, health care consumption and socio-demographic characteristics. By conducting health surveys at frequent intervals, patterns and trends can be spotted and acted upon. Health surveys can be general to capture general health indicators of the general population or disease-specific; focused on a particular disease or condition, such as obesity, cancer recovery or diabetes. In Belgium, many initiatives exist regarding the collection of health data using surveys. One of the larges is the Belgian Health Interview Survey, a cross-sectional population survey that was first carried out in 1997 and has been repeated five times since then. In 2018, the sixth HIS was be organized.

In addition to survey data, objective clinical measures are needed to measure objective health (indicators) across the lifespan. The collection of objective health measures at population level is particularly important for items that are often not accurately reported in a health interview survey and for diseases or health conditions that may be present without any symptoms but have important health consequences. Combining subjective as well as objective health information provides important information on the path to disease or other health outcome, allows to evaluate markers for early detection of disease and identifies individuals at increased risk for developing (chronic) diseases.

1.2. SOCIO-ECONOMIC FACTORS

Identifying the factors contributing to sustained labor force participation and decreasing inequalities is of eminent importance and the Belgian government has identified labor force participation as an important policy objective and dedicated substantial resources, aiming to reduce poverty by increasing the participation rate. Unemployment is a major concern, in particular for the low educated (Cockx, 2013) and large geographical heterogeneity exists (Bodart, 2018). To thoroughly map the barriers to labor force participation and the direction of the causal relationship between social economic disparities (in income, education, and occupation) and health, information from different sources (statistical databases, health registers and lifestyle information) should be linked together. While there is some administrative data on social security and education available, this information lacks qualitative data on the nature of and the interaction between work/employment, well-being, social support, lifestyle factors, etc. In addition, information on the relationship between poverty and migration in a longitudinal design is missing.

1.3. UPBRINDING AND HOUSEHOLD ENVIRONMENT

All these factors are impacted by upbringing and family environment. The characteristics of the living and family environment and the household composition of an individual may have an important effect on the health status. Over the last decades, the composition of Belgium households changed from mainly nuclear families to more non marital cohabitation or consensual unions, one-parent households, single person households etc. Although census data provide us with an overview of the composition of the family, other household related variables are not available, whereas parental milieu and early childhood setting may play a large role in individual's health and social and economic situation in adult life. In addition, data on (the effect of) the decisions that are made within the family context, the care for the child or the parent, etc. currently not available (on a large scale) and many surveys are focused on one individual of a household. Finally, in the analysis of life course of individuals, including the causal effects of major life events or changes, data on family environment is essential.

1.4. 'OMICS'

Finally, an individual's health and social and economic situation are the result of a complex interaction of environmental factors and genetic predispositions. The emergence of advanced methods for analyzing the human genome makes it possible to investigate the understanding and complexity of the influence of genetic factors on health. The use of genomics information to benefit public health is an exciting new field of study. Preventive care and disease treatments maybe more effective by tailoring these to the genetic makeup of an individual. Whole genome sequencing (WGS) of a representative sample of the total population makes it possible to determine genetic variability in the population. However, most of the DNA samples are currently collected and analyzed in clinical populations in a hospital setting. The BelPHG-21 study was a first initiative to describe the genetic variability in the Belgium population using a genome-wide SNP array (Van den Eynden et al., 2018). Nowadays, population-based WGS initiatives, which cover the whole genome, have been performed at the national level throughout Europe to relate the population variability to several indicators of health and disease. To determine the genetic variability in the Belgium population and to gain sufficient power for these study designs, population-based genetic reference data are required.

2. Identifying the correlations

The greatest challenge is to identify the correlations between the factors in Figure 1. Linking data from these different domains on an individual level is crucial. For example, labor force participation is an important policy objective and substantial resources are made available by the Federal Government, aiming to reduce poverty by increasing the participation rate. There are important barriers and resources that have an impact on labor participation, including health and disability, geographical differences, educational level and household characteristics as well as wellbeing and coping, social support, self-efficacy (Augner, 2018; Hergenrather et al., 2015; Zenger et al., 2013). To assess their prospective impact on sustained labor force participation and decreasing inequalities, data on these topics should be brought together to explore their complex interactions. Creating a structural administrative data linkage platform that simultaneously foresees solutions towards retrospective or even prospective data collection allows for unraveling the complex determinants of labor force participation. Another example

entails the understanding of the barriers to seeking health care and mapping of the unmet needs for health care, which is imperative for developing rational policy to provide services that are accessible, efficient, and (cost-)effective. Addressing these barriers include socio-demographic aspects, previous health care experiences of the patients, physical accessibility, but also family environment and social support (EU Expert Panel on effective ways of investing in health, 2016). Here, data obtained in administrative processes such as the social security system is vital to complement (longitudinal) health survey data and when it comes to identifying and direct and indirect effects.

Linking population based register data overcomes limitations in survey data, such as limited data quality due to recall bias and underrepresentation of certain subpopulations. Other countries have built databases or pipelines to facilitate secondary data linkage. Like Belgium, Denmark has many databases and registers containing health data. The Copenhagen Healthtech cluster aims to reduce the administration connected to identification and application for access to Danish health data by providing an overview of the many Danish health data and descriptions of how to apply for access to this data (www.danishhealthdata.com). Norway has a well-developed and well-documented system for register data, Microdata.no, which is developed and operated by the Norwegian Centre for Research Data and Statistics Norway, and provides researchers at approved research institutions access to register data (www.microdata.no). Researchers can use non-anonymized data through a platform that protects confidentiality. In Finland, a one-stop shop for the secondary use of health and social data was established to improve the data security and maximize the use these data: Findata (www.findata.fi). Findata operates in conjunction with Finland's National Institute for Health and Welfare and permits secondary use of data and provides access to anonymized or pseudonymized data for the use in a remote access system for a fixed period of time.

Belgium has many registers and these data sources contain valuable information on population structure, mortality, social economic data, hospital data, etc. and linkage of data is possible due to the existence of the national registration number. Currently, data linkage is mainly project-based, implying that researchers spend a lot of time on inventorying what data is available, requesting the desired data, and linking the data. Sometimes months or even years are spent on reaching consensus with all the data holders, adapting the datasets for researchers and building the pipelines for the data exchange etc. In addition, pooling data often requires the stipulation of a specific timeframe, which hampers longitudinal studies. Routinely linked data or pipelines for fast data exchange, might save valuable time and resources of researchers, and might also aid in supporting individuals working in policy and administration by proving accurate numbers on incidence and prevalence, causes and prognosis with a minimum time lag. Data within closed users networks might not always be easily accessible for third parties. Incorporating this data into a research infrastructure can aid in maximizing the use of these types of data. In addition, it is significantly cheaper to invest in high-quality (secondary) data analyses on existing data than to initiate new studies. Previous reports, such as the EU Expert Group on Health Systems Performance Assessment (HSPA), 2019) already demonstrated repeatedly the need for a more integrated system.

3. Availability longitudinal (health) data

To assess exposure-outcome relations, longitudinal data is essential. Humans are subject to constant change and to document the stable patterns and the changes over time, and to identify new trends longitudinal data is vital. Official statistics or administrative data cannot be a substitute for population-based longitudinal studies. These types of data do not always reflect current scientific issues or issues that are important to assess health issues or health policies.

The strength of combining a longitudinal design with a large sample size is that causality can be explored, taking into account a wide range of determinants in many fields: genomics, lifestyle, social, occupational and physical environment, etc. A cohort can map longitudinally exposure-outcome relationships, as the cumulative interactions between biological, behavioral and sociocultural influences can be studied. Knowledge on how one evolves from health to disease allows to mark risk factors that might be amendable for intervention, which results in a cost-saving effect on health care. In addition, a large sample size allows nested studies in specific subgroups of the population.

4. Belgium-specific (health) policies and interventions

Finally, data from longitudinal studies could be used for successful intervention studies targeting the determinants of health and disease, thereby illustrating the health-policy significance of population-based cohorts. New intervention instruments should be evidence-based, demonstrating the need of the availability of quantitative, standardized data of risk factors for (chronic) diseases and their subclinical manifestations. These data measured over a long observation period allow for an assessment of the influence that individual risk factors (or specific combinations of risk factors) have on the emergence of major widespread (chronic) diseases. By doing so, the appropriate intervention measures of health policies can be developed and implemented, aiming to reduce the incidence rate of the disease.

In order to design a successful intervention or health guidelines or policies, the results provided by population-based longitudinal studies are key, as the strategic aim of health policies is the promotion of preventive approaches in order to reduce health care costs. A good example is the finding of the relationship between level of education/income and the incidence of chronic diseases and premature death in European countries; the lower the social position, the higher the risk (Gallo et al., 2012; Mackenbach et al., 2008). By acknowledging this social gradient of morbidity, the appropriate and policy or intervention can be developed for target groups. Also mapping the determinants of cardiovascular disease (irrespective of the ageing population and population growth) by cohort studies such as the Framingham Heart Study (Tsao & Vasan, 2015) allowed for successful intervention studies such as the North Karelia Study in Finland, which was started in 1972 as a national pilot and demonstration program for CVD prevention (Vartiainen, 2018).

Vice versa, a population-based cohort is the optimal design for evaluation and assessment tool for existing health policies and interventions. The effects of a newly implemented health policy or health guidelines can be explored as longitudinal health data is available within the cohort design and labor policies can only be evaluated in a longitudinal design, as changes in working conditions or health occur with a delay. For example, early retirement is the direct outcome of pension schemes. In countries with a higher retirement age, negative age-related cognitive developments occur later (Börsch-Supan, 2013), which illustrates a clear link with successful ageing.

However, as there are differences between Western-European countries, data collected within cohorts abroad might not be generalizable to Belgium, which stresses the need for national data collecting efforts. Collecting the same information in a Belgium cohort as in cohorts abroad might strike as counterintuitive and redundant, but can aid in mapping the differences in health (policies) in European countries and can support the cost-efficiency of the cohort, as data can be used by researchers abroad. The use of standardized methods can be justified by the need of comparability and increase of statistical power. A large cohort in Belgium might be key in supporting the evaluation of Belgium (health) policies, not only on the federal level, but also on the level of the different regions in Belgium.

5. Position in the European landscape

As Europe faces health inequality within and between Member States, Pan-European comparative studies addressing the impact of underlying determinants (lifestyles, environmental, health(care) systems) of population health and provide new solutions and interventions are needed. There is a strong urge to integrate health information systems, both within and between countries. Countries have to link multiple data sources to provide accurate and complete estimates, for example in upcoming Questionnaires on Diagnosis-Based Morbidity Statistics. The scattered Belgian health information system however makes it very cumbersome for Belgium to fulfill these data requirements. Furthermore, there is a demand at the European level for the establishment of national nodes, which are organizational entities, often linked to a national institution or governmental unit that function as a national liaison and bring together relevant national stakeholders in the country in a systematic way (Joint Action on Health Information, <https://www.inf-act.eu/>).

Additionally, an ESFRI proposal (European Strategy Forum on Research Infrastructures) is under preparation by Sciensano and the consortium DIPoH (Distributed Infrastructure on Population Health). DIPoH is a network of pan-European research groups and national population data curators making this research possible by improving the identification, the access, the assessment and the reuse of individual level and aggregated level population health data and information (data on health status, health determinants (lifestyle, genomic profile, work, socio-economic, environment) and health care data). Improving reuse and linking of different data sources allows researchers, in interdisciplinary teams, to generate larger and powerful datasets, which provide new insights into the dynamics of population health, into the most important influences on health and care, and into the safety, quality, effectiveness and costs of interventions.

6. The case of COVID

In December 2019, a COVID-19 outbreak that started in Wuhan (Hubei, China) December 2019, was identified to be caused by a novel coronavirus, namely severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (World Health Organization, 2020). The outbreak spread across the world in an unprecedented manner and by mid-March, Europe became the epicenter of the epidemic, reporting over 40% of globally confirmed cases. COVID-19 has a large impact on the population, with physical and social distancing becoming the new norm and, where needed, quarantining and self-isolation. As a result, this pandemic has proven to be more than a health crisis; it is a human, economic and social crisis, impacting people, societies and economies at their core. This highlights science's vital role in this crisis: The scientific community has been brought to the forefront as key stakeholders within and outside the health system and policymakers turn to experts looking for answers. This emphasizes the need for research collaborations, the fast availability of data and exchanging scientific knowledge and knowhow.

The exchange of, and access, to up-to-date health data is important for rapid response to research questions that concern current public health challenges, such as the current COVID-19 crisis. In the aftermath of the crisis, information is needed on social, economic, cultural and physical environment (including behavioral data, food, education, work environment, sickness leave/return to work, housing, school) in order to study the effects on health status and population health. A national research infrastructure can facilitate structural data exchange, as well as best practices, methodologies and expertise and allows researchers to provide relevant and evidence based information ready for use in research, and decision-making processes by citizens, clinicians, public health practitioners and policy makers.

OPPORTUNITIES IN BELGIUM

1. National Health Data Platform

1.1. BACKGROUND

Public health challenges such as the ageing population, the emergence of chronic diseases and the identification of risk factors for the development of other disabilities call for new ways of using our health information systems. In order to offer both the political system and scientists a new tool to respond to new public health challenges, linking data from these different domains on an individual level is crucial (see also 2.2). Several high-quality data sources exist in Belgium, such as the mandatory health insurance data, the hospital discharge data, and causes of death data, which provide valuable routine information on the health status of the Belgian population. In addition, data collected in health surveys, cohorts, registers, etc. are managed by a variety of academic and governmental actors, including Sciensano, the Belgian institute for health (Table 3). The setting up of linkages of register data elsewhere in Europe depends on having high quality administrative registers, the presence of personal identifiers and national legislations that approve data linkages (such as Scandinavian countries and The Netherlands).

The Intermutualistic Agency (IMA/AIM) governs a large collection of health care data, as processed by the health insurance services in Belgium and aims to support the improvement of the performance, the quality and the accessibility of the Belgian Health Care system and health/invalidity insurance. In addition, IMA/AIM facilitates the Permanent Sample (Echantillon Permanente Steekproef; EPS), which contains a randomly sampled cohort of health insured individuals in Belgium. This cohort is considered to be a representative sample of the whole of the Belgian population. The database consists of three types of data, one regarding the population, one containing a database with reimbursed health care procedures and one with the reimbursed medication. However, health insurance data have some shortcomings as they provide essentially information on the use of health care and health care expenditures, but information on non-reimbursed health care use or health care needs is lacking. Vice versa, collection of information via health interview surveys such as the BHIS has its limitations such as non-response or selection bias and reporting bias, or information on for example, health care use and expenditure is difficult to estimate using a questionnaire. Individual linkage between the BHIS and health insurance data allows for exploration of these issues. In 2008, the first linkage of BHIS data and Health Insurance data governed by IMA/AIM was realized and confirmed the feasibility of the linkage process. Several studies show the added value of this data linkage (Mimilidis et al., 2014). One of the recommendations of this HISLINK 2008-project was the need to establish a systematic coupling of data. Currently, in the HISLINK 2013-project, data of the BHIS 2013 is linked with IMA data five years before and after the BHIS 2013 as well as the 2013 IMA data. Furthermore, the linkage with the BHIS 2018 is ongoing. This allows for longitudinal analyses of the data, which can provide for example predictors of

nursing-home entry for elderly and an estimation of the evolution of health care consumption and health costs associated with certain diseases or behaviors.

Furthermore, there is the National Hospital Stay Database (NHSD), a merged database that is created by the TCT (Technische Cel – Cellule Technique) from RIZIV/INAMI, to generate a comprehensive overview on the required care, the financial needs and the reimbursed costs, for each specific pathology and to be able to list these costs for each hospital individually and compare the data between all Belgian hospitals. The NHSD is based on the Hospital Discharge Data (Minimal Hospital Data, MHD) and the Hospital Billing Data (HBD). MHD is an administrative data source governed by the Ministry of Health in which (anonymized) medical and nursing data of the (non-psychiatric) hospitals is collected to support the governance and health policies of the Ministry of Health and the hospital (networks). The HBD from the National Health Insurance companies contains the information regarding the billing data for hospitalized patients.

Regional initiatives involve Intego, an integrated network of general practitioners (GPs) in Flanders, managed by the Academic Centre for General Medicine from KU Leuven. Intego aims to create a large database to centralize health data in primary care. The network incorporates an automated data collection, based on the Electronic Medical Record (EMR) of the patient as registered in their GP's practice.

Other databases contain valuable information in relation to health. Statbel, the Belgian statistical office, collects, produces and disseminates reliable and relevant figures on the Belgian economy, society and territory. The collecting is based on administrative data sources such as population structure, movements, births, mortality and migration, as well as surveys, such as data on the Labor Market (Labor Force Survey), Education and training (Survey for the workforce) and poverty and living conditions (form the SILC-survey Statistics on Income and Living Conditions). Another example is the data warehouse of Labor and Social security (governed by the Crossroads Bank for Social Security, CBSS) that aims to link socio-economic data from Belgian social security and other public institutions. At the request of a number of social security institutions, a data warehouse was created in 1999 in which the administrative data is permanently stored. From 2003 onwards, the number of participating institutions was greatly expanded and some sources were also supplemented with additional information. The added value of this data warehouse lies in the linkage of the data of the participating social security institutions.

Finally, Sciensano is responsible for managing some of the major national health surveys, such as the Belgian Health Interview Survey (BHIS) and the Food Consumption Survey (FCS) and surveillance systems. It hosts different disease registries, such as the Belgian Neuromuscular Diseases Registry (BNMDR), Belgian Cystic Fibrosis Registry (BCFR), The Initiative for Quality improvement and Epidemiology in Diabetes (IQED), the network of General Practitioners, coordinates the Belgian national burden of disease study, chairs the European Burden of Disease Network (COST Action CA18218) and coordinates the Joint Action on Health Information (InfAct). In addition, Sciensano has developed a

technical platform for data linkage and transfer, Healthdata.be, that aims to facilitate the data exchange between healthcare professionals and researchers to increase public health knowledge in Belgium.

Table 3. Data holders of health or health related data in Belgium

Data holder	Data
IMA/AIM	Billing data collected by all illness funds
FOD – SPF Public Health	Hospital discharge data and Psychiatric hospital discharge data Workforce Register: national register on healthcare professionals, containing information on new graduates and professionals licensed to practice.
Statistics Belgium	Main statistical authority in Belgium: population and mortality data
RIZIV – INAMI	Farmanet/Pharmanet: database containing information (use, volume, etc.) on all reimbursed medicines in public pharmacies. Information on providers of care and use of health services
Sciensano	Major national health surveys, Belgian national burden of disease study, healthdata.be, surveillance network, disease registries etc.
Belgian Cancer Registry	National register of cancer cases
Intego	Integrated network of general practitioners in Flanders
Crossroads Bank for Social Security	Data warehouse of Labor and Social security: socio-economic data from Belgian social security and other public institutions

However, these different sets of individual health data remain largely unconnected and unharmonized. Currently, data linkage is mainly project-based, implying that researchers spend a lot of time on inventorying what data is available, requesting the desired data, and linking the data. Sometimes months or even years are spent on reaching consensus with all the data holders, adapting the datasets for researchers and building the pipelines for the data exchange etc. In addition, pooling data often requires the stipulation of a specific timeframe, which hampers longitudinal studies. Routinely linked data or pipelines for fast data exchange, might save valuable time and resources of researchers, and might also aid in supporting individuals working in policy and administration by providing accurate numbers on incidence and prevalence, causes and prognosis with a minimum time lag. In addition, it is significantly cheaper to invest in high-quality (secondary) data analyses on existing data than to initiate new studies. Previous reports (e.g. Inventaris van databanken gezondheidszorg, KCE, 2006) show that making an inventory and acquiring data for research is hampered by the fact that the exact characteristics of the databases cannot be obtained in a user-friendly manner. To some extent, this has been addressed by the development of the FAIR Healthdata portal (<https://fair.healthdata.be/>), which aims to provide an exhaustive overview of existing health databases.

To build prospective linkage or pipelines to smoothen the data transition, there is a need to explore and document the possibilities and challenges for setting up an accessible administrative health data cohort in Belgium.

1.2. CONCEPT

To address this need, a conceptual data model in which objectives are set and health and health care data are organized could serve as a starting point for a national health data platform. Based on this conceptual model the role of the responsible organizations for providing the datasets and the procedures and methods to exploit data can be drawn up. The data that are collected within the large health surveys that are managed by Sciensano, such as the BHIS and the FCP, might constitute a starting point for the construction of data transfer pipelines with Belgium's administrative registers, such as Statbel, IMA/AIM, RIZIV/INAMI, aiming for structural pipelines for prospective linkage of data the BHIS and/or FCS with other, administrative sources within the Belgian health information system. In this scenario, the participants of the BHIS/FCS are followed up over time by refreshing their data every year through linking their survey data to the data from the registers.

The BHIS is carried out every 4 to 5 years (Stefaan Demarest et al., 2013; Van Oyen et al., 1997). Each wave comprises of about 10.000 surveyed individuals in some 6.000 households. The survey is commissioned and co-financed by the “Commission of Commissioners”, consisting of the cabinets and administrations of the Ministries and organized by Sciensano. The BHIS commenced in 1997 and was reorganized in 2001, 2004, 2008, 2013, and 2018. The purpose of the BHIS is to monitor the health status of the general population as well as health determinants including health behaviors, medical care consumption and social and demographic characteristics (for domains of the BHIS, see Table 4). The repeated cross-sectional design of the BHIS enables the assessment of health trends and provides evidence for the evaluation of health policy. The target population of the BHIS consists of all persons with residence in Belgium, including the institutionalized elderly, with no restrictions on age or nationality. The National Population Registry (NPR) is used as the sampling frame. This registry contains information on gender, age, address, citizenship, marital status, etc. of each individual. Data collection is undertaken using face-to-face interviews at the participant’s home. The interviews are supplemented with a self-administered questionnaire (for the participants aged 15+) covering more sensitive topics like mental health, use of illicit drugs and sexual behavior. Another large survey effort is the Food Consumption Survey (FCS). The first Food Consumption Survey was conducted in 2004 among 3.249 adults aged 15 and above. In 2014-2015, data were collected for the age groups between 3 and 64 years, with a total sample size of 3461. The participants provided detailed information about their diets via diet survey. In addition, they gave information about their eating habits, their opinions on the food policy, their knowledge of food safety, physical exercise and their sedentary behavior. Weight, height, and waist circumference were also measured objectively. In children and adolescents, physical exercise was measured objectively using an accelerometer.

Table 4. Domains and measurements of the HIS/HES

Domains of the HIS	Additional HES (2018) measurements
Health and wellbeing	Physical examination
Health determinants and lifestyle	Blood sample
Health and social services	Urine sample
Physical and social environment	

Whereas the data holders listed in 3.1.1 have long standing history of data linkage have built up a large expertise and know-how on managing large scale health data, linking the BHIS data to register data was mostly project-based. To scale these experiences up into a more structural, longitudinal cohort, a lot of challenges have to be addressed. In order to build capacity, a thorough identification of the technical bottlenecks and development of an appropriate data management strategy is needed. The current legal framework allows for project-based data linkage between different data sources. However, the possibilities of linking these health data on a more structural basis within the limits of the current legal framework should be identified. This includes a longitudinal commitment of data holders, intellectual property issues, clear rules for engagement for partnerships and solid procedures to draw samples from this data and share the data, all according to the FAIR data principles. Finally, the ethical issues regarding the ethical responsibility of systemic data collection, management and analysis of health data should be explored and addressed, protecting participant and patient rights needs to be balanced with the benefits that big data and linked data collections bring to public health research. The fundamental ethical principle of autonomy demands that the wishes of all participants must be respected by asking their opinion or willingness to get involved or not – as stipulated by current legal frameworks. However, epidemiological research can involve the use of participants' data without their consent, resulting in a loss of autonomy and a need for other ethical justification.

In order to understand the dynamics of public health, linking different sources of administrative data is not sufficient to perform large-scale epidemiological studies or health policy evaluations. The availability of longitudinal data is essential for these purposes. In this opportunity, the participants are followed up over time by refreshing their data every year through linking their survey data to the data from the registers. Other large population based data collection in Europe, such as the UK Biobank, also conduct follow-up for health-related outcomes through linkages to routinely available national datasets such as hospital admissions, mortality, cancer registries, primary care, cancer screening data, and disease-specific registers (Sudlow et al., 2015). Another example is the Estonian Biobank, where the follow-up is also conducted by enrichment of phenotypic data by updating the health data from medical records available in national health databases and hospital registries (Leitsalu et al., 2015). This initiative will include tens of thousands of participants with data on health and lifestyle, as well as administrative data on demography, social security, education and health care, and provides a research infrastructure where data can be accessed by researchers, administrators and policymakers.

As a thorough identification of these bottlenecks is of eminent importance, a recent project proposal was drawn up (AHEAD project) in order to address the lack of an integrated health information system in Belgium, by exploring the possibility of harmonizing health data that exist within the Belgian health information system and making these accessible for scientific exploitation and scientific valorization. Ultimately, this existing research infrastructure can be the foundation for, and be developed into, a cohort study, in which (a) (new) wave(s) of cohort participants can be followed over a longer period of time.

1.3. IMPACT & RISKS

Such an infrastructure impacts scientific and technical knowledge developed with respect to the Belgian health information system and the possibilities to develop a national health data platform. It will serve a variety of users, of which the primary users are researchers in public health and population sciences or other connected fields. The infrastructure will enable researchers and their communities to perform excellent cross-disciplinary and data-intensive research, which will improve the use and valorization of the data available within the Belgian health information system, and in turn will support public health policy and scientific research. Through such a platform, the academic community will be able to access an overview of the Belgian health information system, and extract the characteristics of the databases per institute or source. By providing information on the modalities for the exploitation of these data (contact persons, application procedures and pipelines), valuable time and resources will be saved, as currently, a lot of time is spent on inventorying and requesting the data needed. In the longer run, such a platform will serve as a novel, powerful research infrastructure for population health research, fostering public health policy and strengthening the research communities in Belgium. Indeed, a national health data platform could aid in underpinning policy decisions by providing accurate numbers on incidence and prevalence, causes and prognosis with a minimum time lag. In addition, such a platform would provide novel opportunities for evaluating existing health policies and interventions. The effects of a newly implemented health policy or health guidelines can be explored as longitudinal health data is available and labor policies can only be evaluated in a longitudinal design, as changes in health occur with a delay. Finally, a large platform comprising health data that is representative of the Belgian population does not only support the evaluation of health policies on the federal level, but also on the level of the different regions in Belgium.

Some risks need to be addressed. A lack of interest or involvement of stakeholders pose a risk as this project depends heavily on interaction with the different actors of the Belgian health information system. Fortunately, Sciensano has a privileged position in the Belgian health landscape and has a broad network of partners and regularly meets with them to support knowledge translation and preliminary discussions with possible stakeholders. Importantly, this project aims to identify the bottlenecks of built around prospective linkages, thus documenting any difficulties in aligning input from different stakeholders is an objective in itself.

The prospective linkage of Sciensano's Health Interview Survey data to the data from the administrative registers will serve as a business case, in order to explore the technical bottlenecks, identify the possibilities within the current legal frameworks while taking into account the ethical aspects. However, by taking the BHIS 2018 as starting point, this initiative cannot compare itself to the larger cohorts listed in Table 1 with regard to sample size. This sample (N~11,000) not representative for larger scale infrastructure. Nevertheless, within the a stratified multistage, clustered sampling method is applied, in which the regional (Flemish Region, Walloon Region, Brussels Capital Region) samples are fixed. Moreover, from a health research perspective it is of eminent importance to have access to survey data to get an overall view of the health of the population. The BHIS is the largest health surveys countrywide

and covers domains such as health and wellbeing, health determinants and lifestyle, health and social services, physical and social environment, and prevention.

Table 5. SWOT-analysis of setting up a national health data platform

Strengths	Weaknesses
<ul style="list-style-type: none"> • Feasibility of coordination of a multi institutional initiative is demonstrated by successful project-based data linkage initiatives • Builds on existing research projects • Involvement of different stakeholders consolidates collaboration and sustainability of a research infrastructure • Provide overview of data and methodologies • Capacity building, best practice exchanges and increased cooperation between data holders and research networks • Pooled resources for more effective health data collection, dissemination and analysis • Mutualized learning and data access 	<ul style="list-style-type: none"> • Focus on optimizing access to population health data for research purposes, but not yet on the implementation into policy and practice. • Business case of HIS is not representative for larger scale infrastructure • Longitudinal data on lifestyle or health behaviors is not available • Epidemiological research can involve the use of participants' data without their consent, resulting in a loss of autonomy and a need for other ethical justification
Opportunities	Threats
<ul style="list-style-type: none"> • Uptake of innovations in the domain of population health data collection, reporting and use for surveillance and research • Short lines of communication between different data holders and stakeholders. • Increased opportunity to ensure FAIRness of population health data • Landscape analysis and inclusion of researchers/stakeholders in other sectors that have an impact on health • Enrichment of available datasets through linkages between the data holders • Setting up a retrospective cohort could be an option • Integrated approach to population health monitoring and health system performance • Closing health information gaps by linking with other European countries, contributing to global data and research initiatives 	<ul style="list-style-type: none"> • Smaller stakeholders that are not part of the initial project do not find sufficient incentives to contribute. • A lack of interest or involvement of stakeholders as this project depends on interaction with the different actors of the Belgian health information system. • Data holders are not convinced that sharing of data is safe and do not want to participate in the process. • No workable solutions for a viable funding model can be found • Deeply differing opinions of important technological and structural choices within the research infrastructure • Scattered competences in domains of health and health related information in the institutional organization of Belgium

2. Online panel

2.1. BACKGROUND

Using online research panels is a quick and cost-effective way to conduct market research. With more and more of our daily lives spent online, shifting from pen-and-paper questionnaires to online surveys has opened up a world of convenience for researchers and allowed them to reach a large number of people in a short period of time. In other European countries, large internet panels are used to collect information on lifestyle, health behavior, diseases and expectations and experiences of health care users. The LISS Panel in the Netherlands (www.lisspanel.nl): the ELIPSS panel in France (www.elipss.fr) and the German Internet Panel (GIP, <https://www.uni-mannheim.de/en/gip/>) in Germany, varying from 1500 individuals to 5000 households are academically-based online panels that are currently operating. Although this presents quite some challenges, such as false answers, careless responses, giving the same answer repeatedly, getting multiple surveys from the same respondent, and panelists being members of multiple panels, this approach is cost-effective, it enables access to large and diverse samples quickly, and can be replicated easily (Hays et al., 2015).

In the context of the European Health Interview Survey (EHIS), many European countries already adopted a web-based mode (as part of a mixed-mode design). The EHIS provides harmonized statistics on the health status, health care use and health determinants of the European Union and European Economic Area populations (Eurostat, 2018) and these general population health surveys serve as an important data source for monitoring population health and policymaking. However, web-only data collection may not be favored over interviews or mixed-mode data collection efforts due to low response rate, especially for some subgroups in the population. In Belgium, lower web response rates were obtained among the elderly, lower educated people, people with a foreign background, people residing in Brussels Capital and/or living without a partner (Braekman et al., 2020). However, also for many of the population-based cohorts listed Table 1 (except for the Swiss National Cohort, which includes the whole population) the existence of a healthy volunteer bias is suggested, as participants were more likely to be female, less likely to be obese and to smoke on a daily basis and have higher socioeconomic status than nonparticipants.

Between April and June 2020, three COVID-19 health surveys were organized, targeting all adult inhabitants of the Belgium (minimum age: 18 years). As the survey is an online survey, persons who did not have an Internet connection were excluded. The recruitment of participants was mainly done via the snowball principle: when the survey was announced, it was asked to forward the invitation to family, friends and acquaintances. The survey and the relevant access link were announced via the Sciensano website and other organizations (mutual health organizations, community centers...), through the press and social media. For the second and third survey, all participants of both (or one of both) previous COVID-19 Health Surveys, who had expressed their willingness to participate in the next survey (and had provided their e-mail address for this purpose) were invited via an e-mail. The sample size ranged from 44.000+ for the first wave, to more than 33.00 for the third wave (Demarest et al., 2020b, 2020a; Drieskens et al., 2020). Taking the third wave as an example, the participants residing in the Flemish

Region were overrepresented and participants in the Walloon Region were underrepresented in the sample. Furthermore, the composition of the sample was very different from the composition of the population (18 years and over) in terms of gender, age group and level of education. In order to minimize this selection bias, the results of the analyses are weighed.

As stated in 2.3, to assess exposure-outcome relations, longitudinal data is essential. Humans are subject to constant change and to document the stable patterns and the changes over time, and to identify news trends longitudinal data is vital. The BHIS exploits a cross-sectional study design, which makes it less optimal for longitudinal data analysis. An online panel allows for longitudinal data collection, as participants can be invited to participate in multiple questionnaires over the years.

2.2. CONCEPT

These experiences demonstrate that the possibility of setting up an internet panel is one to consider. There are several ways to recruit participants for the online panel. Recruitment is possible through online self-registration: for large cohorts abroad, such as LifeLines (NL), 13% of the participants self-registered via the LifeLines website (while 49% of the included participants was recruited through their GP, and 38% were family members of the invited individuals). However, most cohorts abroad aim to be representative of the targeted population by creating sampling frames created by their national registries and participants are invited to include their household members. As information on the family environment and the household structure of an individual is very important, whole households should be included as sampling unit, including children and subjects that are yet to be born within these household. By doing so, short term research goals, such as providing information on health behaviors in adulthood can be achieved, as well as long term goals, as the foundation of health is already being established in childhood.

In Belgium, the National Population Registry can be used as sampling frame, as is the case for the BHIS. Stratified multistage clustered sampling can be used to compose the sample: a number of municipalities and households can be selected in which the survey will be conducted. However, initial response rate in the web version of the EHIS in other countries is low (Braekman et al., 2020) and the 'snowball principle' showed (albeit under specific circumstances) for a large support base and therefore a large sample size, although a carefully designed and implemented online panel can produce high-quality data at lower marginal costs (Blom et al., 2015). For the three EU panels listed in 3.2.1, all participants were selected by the national or regional statistical offices, and analyses into the representativeness of the online panels showed no major coverage or nonresponse biases. The biggest challenge regarding response rates might not lie in the follow-ups, but in the first wave of data collection.

The second phase of the panel, involves the longitudinal data collection. This internet panel can evolve into a cohort, when data collection is repeated over time. The strength of combining a longitudinal design with a large sample size is that causality can be explored. A longitudinal data collection effort can map longitudinally exposure-outcome relationships, and thereby mark risk factors that might be amendable for intervention. Many population based cohorts collect data in waves: the average time between the

waves is 5 years, but internet panel provide the flexibility to send questionnaires at shorter intervals, for example every 2 or 3 years. The European Social Survey (ESS) might provide a good example: it consists of a collection of questions that can be classified into two main parts – a core section and a rotating section. The core section (also referred to as the ‘core module’) focus on a range of different themes that are largely the same in each round. The rotating section (also known as ‘rotating modules’) is dedicated to specific themes, which are sometimes repeated in later rounds of the ESS. This flexibility is a worthy selling point: not only can it provide the opportunity to answer the current burning questions, it might also appeal to future policy makers as they can also reap the benefits from this infrastructure by adding new modules.

To facilitate interdisciplinary research, opportunities should be sought to link data of the participants to register data, as described in 3.1.2. A data linkage can aid in reducing the survey length: sections on socio-demographics, diseases or chronic conditions or the use of health care can be shortened. Shorter questionnaires reduce the burden on participants and could lead to higher response rates (Galesic & Bosnjak, 2009). In addition, it allows to assess response bias and to compare participants and non-participants.

Ultimately, this panel can mature, and additional data collection in a clinical or experimental setting can be possible. A good example is the Netherland Twin Register; In 1986 researcher began this register by recruiting young twins and multiples a few weeks or months after birth, by sending out questionnaires. Nowadays, subgroups of twins and siblings take part in studies of cognitive development, brain function and neuropsychological indices of attention processes, and molecular genetic studies of classical and behavioral cardiovascular risk factors (Boomsma et al., 2006). This requires investments in engagement of the participants; to keep the ‘cohort alive’ time and effort should be invested in keeping the attention of the participants, for example by sending newsletters, the use of social media and other incentives. Furthermore, a feedback mechanism in the form an online dashboard, where participants can log onto and find their personal data in relation to the mean in the overall panel (Willemsen et al., 2014), could serve as an incentive to participate. Finally, participants can be invited in focus groups: small, but demographically diverse group of people who are guided or open discussions about the panel and their participation. Ultimately, when adherence is low and drop-out of participants is high, new groups of participants should be included (refreshment sample), taking the European SHARE study as an example: respondents who were interviewed in any previous wave are part of the longitudinal sample, ensuring a longitudinal design and refreshment samples are drawn regularly to compensate for the reduction in panel sample size due to attrition and maintain representation of the younger age-cohorts of the target population. Figure 2 demonstrates the evolution from such a survey to a more cohort-like initiative.

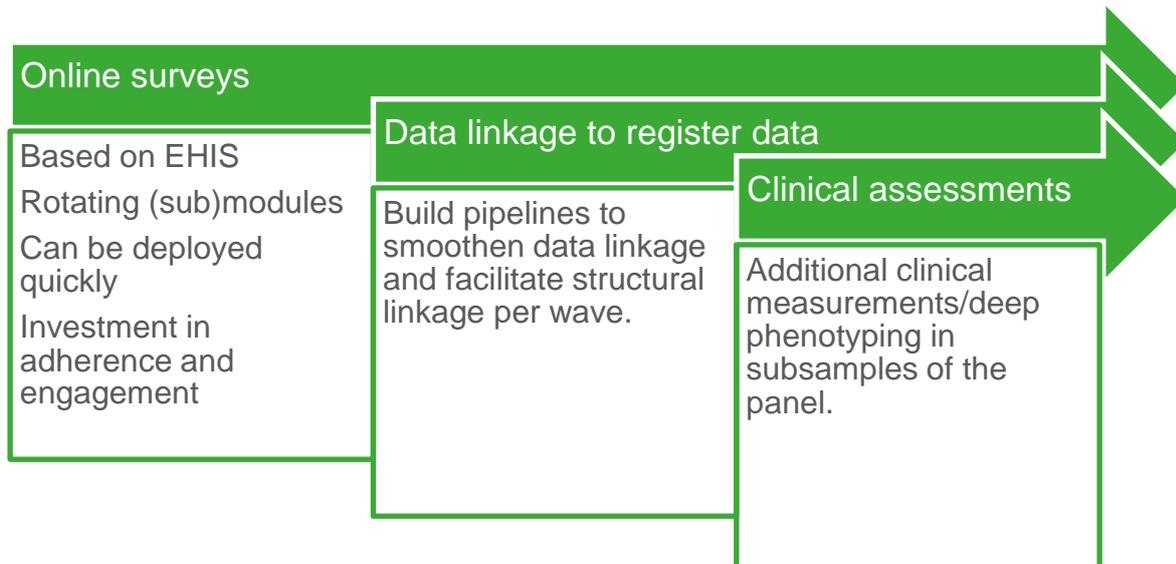


Figure 2. From online panel to cohort.

More support will be found if there is a push policy which promotes data accessibility. To enhance user-friendliness, the information on the design and data content of surveys should be consistently processed and made centrally available and thus centrally managed. Previous experiences show that data is requested by and shared with third parties only when it is free and easily accessible. A solid procedure should be put in place to address the rules of engagement and to regulate and authorize requests to draw samples and share the data, taking into account the different levels of data; for example, biological samples stored in biobanks should be less accessible.

2.3. IMPACT & RISKS

An online panel is a low cost, easy to implement tool to gather data on population health, allowing to reach a large number of people in a short period of time. A pilot study in a sample of ~1000 participants for a web version of the HIS showed that the total cost per completed questionnaire was almost three times lower (€41) compared to the face-to-face data collection in the regular waves of the BHIS (€111). The numerous questionnaires that were initiated at the start of the COVID-19 health crisis show that it is a tool that can be deployed very quickly. This promote rapid data collection analyses of population health including health status, health determinants and health system performance.

Furthermore, as a longitudinal data collection initiative maps longitudinally exposure-outcome relationships, and thereby mark risk factors that might be amendable for intervention, it can be cost-efficient. Although it is difficult to provide an estimation, establishing such a population-based cohort might be more cost-effective in the long run (compared to a scenario in which Belgian cohort-data is not available). A cohort might reduce the costs of maintaining small local initiatives, as these can be incorporated in the infrastructure.

Possible risks include a low response rate, especially for some subgroups in the population and an underrepresentation of these subgroups in the net sample. As a result, possible conclusions drawn from the analysis of results can be biased. However, to compensate for this, post-stratification weights can be used in the analysis. This technique consists of using information about the composition of the population (in terms of gender, age group, province and education) from other data sources to be corrected when calculating the results of the survey. Furthermore, efforts to increase the response and reduce bias such as by incorporating adaptive survey designs (i.e. assigning different survey strategies to different population units) should be further explored. Finally, existing surveys such as the EHIS can be tailored towards an optimized web-based mode: the questions and instructions could be altered in order to give the same perceived stimuli to respondents as questions posed through face-to-face interviewing, such as warning messages in case of inconsistent answers as well as motivational phrases and feedback mechanisms (DeLeeuw, 2018).

Enriching the participants in an online panel with clinical data and/or subjective data might present some complications. When a specific subsample of an administrative dataset is enriched, one should be wary of 'over-surveying' this more deeply phenotyped subsample in other studies. In addition, by providing the opportunity to enrich the data with clinical examinations, subjective data etc. in order to be considered a real asset, consent should be provided by the subjects in the sample.

Such an endeavor requires a long term investment and commitment. Stable long-term funding increases the value of longitudinal data exponentially with the number of study waves. Currently, funding for scientific research comes in smaller (sometimes personal) grants, which are awarded in competitive application processes. This results in an inability to allow for long-term research goals and structural data management, which is defining in creating a sustainable cohort. Large cohorts abroad list their federal government (and/or specific ministries) as their main source of funding, together with national research councils and academic institutes. UK Biobank also lists charity organization such as Diabetes UK and the British Heart Foundation. Although the HUNT has largely been publicly funded, funding is also being provided by commercial enterprises, such as Statoil, and GlaxoSmithKline.

Table 6. SWOT-analysis of an online panel as startup for a national cohort

Strengths	Weaknesses
<ul style="list-style-type: none"> • Low cost, easy to implement tool • Can be deployed quickly when needed • Builds on existing research projects, and the pilot studies on the EHIS in Belgium • The large sample size allows for nested studies • Possibility to add a refreshment sample when drop-out is high 	<ul style="list-style-type: none"> • Focus on subjective data collection, but not yet on the implementation into policy and practice. • It is difficult to keep the 'cohort alive' in terms of adherence and engagement. • A lack of interest or involvement of stakeholders, rendering the (two-yearly) data linkage infeasible. • Potential 'over-surveying' of subsamples • Not yet a legal framework to enrich data with secondary data or clinical assessments in a structural way

Opportunities	Threats
<ul style="list-style-type: none"> • Landscape analysis and inclusion of researchers/stakeholders in other sectors that have an impact on health • Opportunities for scaling up of the data collection, including clinical assessments • Enrichment of the sample through linkages with secondary data holders • Quality improvements in the use of composite health indicators, as longitudinal data is available 	<ul style="list-style-type: none"> • Drop out is high, response rate low, resulting in a non-representative online panel. • No workable solutions for a viable funding model can be found • Deeply differing opinions on what questions and modules to add to the survey waves. • No long term funding opportunities

3. Other Opportunities

3.1. MULTIPLE INITIATIVES, ONE PORTAL

Currently, many data holders of health data exist within Belgium, of which the larger ones are listed in 3.1.2. Data collection efforts that exist for example within an academic environment, often linked to individual research groups with a closed user's network, or in a hospital environment (patient cohorts). In addition, other smaller institutes have also collected health data or related variables, such as the ENVIRONAGE Birth Cohort Study (Janssen et al., 2017), the IPANEMA study (Van den Eeden et al., 2018), or cohorts with elderly; the SarcoPhAge study (Beaudart et al., 2015), the SENIOR cohort (Buckinx et al., 2016) and many more.

It is significantly cheaper to invest in high-quality (secondary) data analyses on existing data than to initiate new studies. However, researchers spend a lot of time on inventorying what data is available in Belgium, requesting the desired data, and linking the data, if necessary. Sometimes months or even years are spent on reaching consensus with a data holder, adapting the datasets for researchers and building the pipelines for the data exchange etc. In addition, data availability requires the stipulation of a specific timeframe, which hampers longitudinal studies.

What is needed, is an online portal providing a clear and concise overview of the key health data sources in Belgium, modalities for the exploitation of these data, their contact persons, and practical information on how linkages can be set up which facilitates researchers to easily find the desired health data collections – which is for instance the aim of the French health data hub (<https://www.health-data-hub.fr/>). A key feature of this platform, is the enabling for the academic community to access an overview studies in health data, and exact characteristics of the databases per institute or source, and links to the FAIR catalogues (Wilkinson et al., 2016). Creating such a platform, might save valuable time and resources of researchers, and might also aid in supporting individuals working in policy and administration by proving accurate numbers on incidence and prevalence, causes and prognosis with a minimum time lag. Healthdata.be's expertise and know-how are indispensable in order to map the difficulties in standardizing and homogenizing the preservation of health data for research purposes.

This expertise can be exploited to create a metadata catalogue to describe population health data sources in Belgium.

3.2. TECHNOLOGICAL ADVANCES IN LARGE SCALE LONGITUDINAL RESEARCH

However, the abundance of administrative data that is available in Belgium might not be sufficient in providing enough information to face the challenges in the domain of public health (see also 2.1.1.) In most of the cohorts listed in Table 1, a questionnaire was administered, as well as clinical examinations which generally included (but are not limited to) anthropometrics, cardiovascular system measurements (such as resting blood pressure and heart rate), spirometry, fitness tests, electrocardiogram and cognitive functions. A subset of the German National Cohort was examined by whole-body MRI. In addition, this cohort as well as the UK Biobank, equipped their participants with an accelerometer to assess weekly physical activity. Most biological samples comprise of blood, which allows analysis of genome wide genotype data, as well as other haematological and biochemical measures.

Advances in ambulatory measurements allow for a new way of collecting data. Ambulatory monitoring is a method of acquiring behavioral and physiological data in subjects who are free to go about their normal daily activities, outside the confines of the laboratory or clinical environment. In the past decades, ambulatory monitoring has evolved from an innovative tool in fundamental research to a widely used method in clinical and applied research settings. Because ambulatory monitoring takes place during everyday life, in the subject's own environment, such measurements have high ecological validity. In addition, it captures physiological processes that have a prolonged time scale, including circadian rhythms and wake-sleep patterns. Ambulatory measurements can include physical variables, such as accelerometry to indicate movement and physical activity, heart rate or other autonomic nervous system responses to indicate physical activities and stress, but also experience sampling methodology (ESM), to capture momentary psychological processes.

Moreover, the emerging field of precision health allows for disease prevention and detection by monitoring health and disease based on an individual's risk: treatments or interventions should be formulated on individuals' unique background and conditions and genomic information can contribute to a better understanding of the individual's health status, risk of developing chronic conditions and sensitivity to treatment. In sample material that can be collected through non-invasive methods, more specifically saliva and urine, biomarkers and (epi)genetic markers can be assessed. Test kits to obtain those types of samples can be sent and returned by mail, without the need of specialized staff required for sample collection. This has already been applied in the framework of the BelPHG-21 study (Van den Eynden et al., 2018). Currently the BIOTECH Platform and Bioinformatics platform of Sciensano is refining methods to extract proteins and DNA from saliva and urine, which will be very useful for the collection of data on biomarkers and genomic information in the framework of population based studies. At the moment, a project proposal for a pilot study is developed in which objective data is collected continuously by apps and smart wearable sensors to examine the potential of wearables in large scale longitudinal evidence-based research. This project, *mPower*, rests on the observation that state-of-the-art mobile sensor technologies (i.e. wearables) in combination with innovative machine learning

techniques hold potential for shortcomings of survey research as well as clinical measurements. In addition, it is examined how predictive performance is improved by genomic information. Here the BHES provides a starting point for a pilot study: In Belgium in 2018, objective health information (a questionnaire, a physical examination and a blood and urine sample) was collected among a representative sample of the total population (N~1100), who also participated in the HIS (Nguyen et al., 2020) by trained nurses. In addition, consent is obtained for linking this data to other databases.

RECOMMENDATIONS

In Belgium, multiple high-quality data sources exist, but an integrated national health information system is lacking; a more sustainable and longitudinal linkage between data sources is missing, hampering the valorization of these data sources. Routinely linking data might save valuable time and resources of researchers, and might also aid in supporting individuals working in policy and administration by providing accurate numbers on incidence and prevalence, causes and prognosis. Nonetheless, different elements already seem to be in place, including the existence of a unique national register number which can be used to link multiple databases, as well as the existence of technical platforms for data linkage and transfer. There is a need for exploring the possibility of harmonizing health data and making these accessible for scientific exploitation and scientific valorization, working towards a national health data platform, thereby integrating the Belgian Health information system. The prospective linkage of Sciensano's Health Interview Survey (HIS) data to the data from the administrative registers can serve as a business case throughout this process.

As a thorough identification of the ethical, legal and technical bottlenecks of such an endeavor bottlenecks is of eminent importance, a recent project proposal was drawn up (AHEAD project, submitted to Belspo June 2020).

Secondly, online research panels are a quick and cost-effective way to conduct market research. With more and more of our daily lives spent online, shifting from pen-and-paper questionnaires to online surveys has opened up a world of convenience for researchers and allowed them to reach a large number of people in a short period of time. In other European countries, large internet panels are used to collect information on lifestyle, health behavior, diseases and expectations and experiences of health care users. In addition, the experiences of the web-based mode of the European Health Interview Survey (EHIS) showed the possibilities for employing this mode of surveying in Belgium. This internet panel can evolve into a cohort, when data collection is repeated over time. The strength of combining a longitudinal design with a large sample size is that causality can be explored. Ultimately, this panel can mature, and additional data collection in a clinical or experimental setting can be possible. The rapid results, the flexibility and the longitudinal perspective, the opportunities for scaling up of the data collection, including clinical assessments and enrichment of the sample through linkages with secondary data holders, makes this tool a worthy backbone for public health research.

At the moment we are investigating the possibilities of translating this into a project proposal; reaping the benefits of the successful COVID-HIS surveys, COVID hospital cohort and recently completed thesis by Braekman (in preparation).

Finally, survey data and register data might not be sufficient; to measure health across the lifespan, objective clinical measures are needed. Advances in ambulatory measurements and wearables allow for a new way of collecting these types of data. Ambulatory monitoring is a method of acquiring behavioral and physiological data in subjects who are free to go about their normal daily activities,

outside the confines of the laboratory or clinical environment. Ambulatory measurements can include physical variables, such as accelerometry (movement/physical activity), heart rate or other autonomic nervous system responses (physical activities/stress), experience sampling methodology (momentary psychological processes), but also material that can be collected through non-invasive methods, more specifically saliva and urine, for the assessment of biomarkers and (epi)genetic markers. The Health Examination Study (HES) can be exploited as use case and pilot study; to examine the possibility of re-inviting participants for follow up, including the use of wearables and genome sequencing.

Currently, a project proposal for a pilot study (mPower) is developed in which objective data is collected continuously by apps and smart wearable sensors to examine the potential of wearables in large scale longitudinal evidence-based research. Submitted to FWO (Oct 2020).

In order to build a strong foundation for a cohort, these opportunities will be further explored, provided funding is obtained. The realization of these opportunities requires a long-term financial and organizational effort from the start. Currently, funding sources for scientific research are split up in smaller grants, which are awarded in competitive application processes, resulting in an inability to allow for long-term research goals and structural data management. Stable long-term funding increases the value of longitudinal data exponentially with the number of study waves. More support will be found if there is a push policy which promotes data accessibility. To enhance user-friendliness, the information on the design and data content should be consistently processed and made centrally available and thus centrally managed. Furthermore, efforts to create a more focused platform targeted at researchers, to providing a clear and concise overview of the key health data sources in Belgium, their contact persons, and practical information on how linkages can be set up should be supported and expanded.

Finally, as Europe faces health inequality within and between Member States, Pan-European comparative studies addressing the impact of underlying determinants of population health and provide new solutions and interventions are needed (see e.g. DIPoH - Distributed Infrastructure on Population Health; and PHIRI - Population Health Information Research Infrastructure). We are duty bound to improve the identification, the access, the assessment and the reuse of individual level and aggregated level population health data and information.

CONCLUSIONS

There is a need for a cohort, organized as an accessible research structure, containing multiple variables, collected within a longitudinal design. The success of initiatives abroad, such as Lifelines in the Netherlands and CONSTANCES in France show the need for such infrastructures.

In Belgium, many opportunities to set up a population-based cohort exist albeit perhaps not in the classical sense of a cohort, such as the LifeLines in the Netherlands or CONSTANCES in France. These opportunities should be explored in depth, consolidating collaboration between different stakeholders and sustainability of the resulting research infrastructure. The current COVID-19 crisis shows that the rapid access to up-to-date health data is of eminent importance. Belgium, being a country with many authorities, services and agencies, but short communication lines, might hold the right cards to build an infrastructure holding longitudinal data collections of register data, survey data and/or clinical data.

Sciensano is responsible for managing some of the major national health surveys, coordinates the Belgian national burden of disease study, leads various projects that involve ad hoc linkages of different health datasets, and hosts the healthdata.be platform for secure collection, transfer and analysis of health data. Moreover, Sciensano is the coordinator of PHIRI and DIPoH. These activities, further validated by its role as coordinator of the Belgian focal point for international data collection, give Sciensano a central role in the Belgian health information system and should be the institute to coordinate and manage this infrastructure. This will strengthen the scientific potential of Sciensano and support its long-term strategic priorities in terms of its public service missions.

REFERENCES

- Almqvist, C., Adami, H.-O., Franks, P. W., Groop, L., Ingelsson, E., Kere, J., Lissner, L., Litton, J.-E., Maeurer, M., Michaëlsson, K., Palmgren, J., Pershagen, G., Ploner, A., Sullivan, P. F., Tybring, G., & Pedersen, N. L. (2011). LifeGene—a large prospective population-based study of global relevance. *European Journal of Epidemiology*, *26*(1), 67–77. <https://doi.org/10.1007/s10654-010-9521-x>
- Augner, C. (2018). Health, education and employment status of Europeans aged 60 to 69 years: results from SHARE Survey. *Industrial Health*, *56*(5), 436–440. <https://doi.org/10.2486/indhealth.2017-0220>
- Beudart, C., Reginster, J. Y., Petermans, J., Gillain, S., Quabron, A., Locquet, M., Slomian, J., Buckinx, F., & Bruyère, O. (2015). Quality of life and physical components linked to sarcopenia: The SarcoPhAge study. *Experimental Gerontology*, *69*, 103–110. <https://doi.org/10.1016/j.exger.2015.05.003>
- Blom, A., Gathmann, C., & Krieger, U. (2015). *Setting Up an Online Panel Representative of the General Population: The German Internet Panel*. *27*(4). <https://doi.org/10.1177/1525822X15574494>
- Bodart, V. D. (2018). The labor market in Belgium, 2000–2016. *IZA World of Labor*. <https://doi.org/10.15185/izawol.428>
- Boomsma, D. I., Geus, E. J. C. de, Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J.-J., Posthuma, D., Beijsterveldt, T. C. E. M. van, Hudziak, J. J., Bartels, M., & Willemsen, G. (2006). Netherlands Twin Register: From Twins to Twin Families. *Twin Research and Human Genetics*, *9*(6), 849–857. <https://doi.org/10.1375/twin.9.6.849>
- Bopp, M., Spoerri, A., Zwahlen, M., Gutzwiller, F., Paccaud, F., Braun-Fahrländer, C., Rougemont, A., & Egger, M. (2009). Cohort Profile: The Swiss National Cohort—a longitudinal study of 6.8 million people. *International Journal of Epidemiology*, *38*(2), 379–384. <https://doi.org/10.1093/ije/dyn042>
- Börsch-Supan, A. (2019). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7* [Data set]. SHARE–ERIC. <https://doi.org/10.6103/SHARE.W7.700>
- Braekman, E., Charafeddine, R., Demarest, S., Drieskens, S., Tafforeau, J., Van der Heyden, J., & Van Hal, G. (2020). Is the European Health Interview Survey online yet? Response and net sample composition of a web-based data collection. *European Journal of Public Health*, *30*(3), 595–601. <https://doi.org/10.1093/eurpub/ckz206>
- Buckinx, F., Reginster, J. Y., Petermans, J., Croisier, J. L., Beudart, C., Brunois, T., & Bruyère, O. (2016). Relationship between frailty, physical performance and quality of life among nursing home residents: the SENIOR cohort. *Aging Clinical and Experimental Research*, *28*(6), 1149–1157. <https://doi.org/10.1007/s40520-016-0616-4>
- Cockx, B. (2013). Youth unemployment in Belgium: diagnosis and key remedies. *INTERECONOMICS*, *48*(4), 202–209. <https://doi.org/10.1007/s10272-013-0464-9>
- CONSTANCES team, Zins, M., & Goldberg, M. (2015). The French CONSTANCES population-based cohort: design, inclusion and follow-up. *European Journal of Epidemiology*, *30*(12), 1317–1328. <https://doi.org/10.1007/s10654-015-0096-4>
- DeLeeuw, E. D. (2018). Mixed-Mode: Past, Present, and Future. *Survey Research Methods*, *Vol 12*, 75–89 Pages. <https://doi.org/10.18148/SRM/2018.V12I2.7402>
- Demarest, S, Braekman, E., Charafeddine, R., Drieskens, S., Gisle, L., Hermans, L., & Scohy, A. (2020). *Derde COVID-19 gezondheidsenquête: eerste resultaten* [Text]. <https://www.sciensano.be/en/biblio/derde-covid-19-gezondheidsenquete-eerste-resultaten>
- Demarest, S, Braekman, E., Charafeddine, R., Drieskens, S., Gisle, L., Hermans, L., & Vandevijvere, S. (2020). *Eerste COVID-19 gezondheidsenquête: eerste resultaten* [Text]. <https://www.sciensano.be/en/biblio/eerste-covid-19-gezondheidsenquete-eerste-resultaten>

- Demarest, Stefaan, Van der Heyden, J., Charafeddine, R., Drieskens, S., Gisle, L., & Tafforeau, J. (2013). Methodological basics and evolution of the Belgian health interview survey 1997–2008. *Archives of Public Health*, *71*(1), 24. <https://doi.org/10.1186/0778-7367-71-24>
- Derom, C., Vlietinck, R., Thiery, E., Leroy, F., Fryns, J.-P., & Derom, R. (2002). The East Flanders Prospective Twin Survey (EFPTS). *Twin Research*, *5*(5), 337–341. <https://doi.org/10.1375/136905202320906057>
- Drieskens, S., Braekman, E., Charafeddine, R., Demarest, S., Gisle, L., & Hermans, L. (2020). *Tweede COVID-19 Gezondheidsenquête: eerste resultaten* [Text]. [sciensano.be](https://www.sciensano.be). <https://www.sciensano.be/en/biblio/tweede-covid-19-gezondheidsenquete-eerste-resultaten>
- Eurostat. (2018). *European Health Interview Survey (EHIS wave 3) methodological manual*. Publication Office of the European Union.
- Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, *73*(2), 349–360. <https://doi.org/10.1093/poq/nfp031>
- Gallo, V., Mackenbach, J. P., Ezzati, M., Menvielle, G., Kunst, A. E., Rohrmann, S., Kaaks, R., Teucher, B., Boeing, H., Bergmann, M. M., Tjønneland, A., Dalton, S. O., Overvad, K., Redondo, M.-L., Agudo, A., Daponte, A., Arriola, L., Navarro, C., Gurrea, A. B., ... Vineis, P. (2012). Social Inequalities and Mortality in Europe – Results from a Large Multi-National Cohort. *PLoS ONE*, *7*(7), e39013. <https://doi.org/10.1371/journal.pone.0039013>
- German National Cohort (GNC) Consortium. (2014). The German National Cohort: aims, study design and organization. *European Journal of Epidemiology*, *29*(5), 371–382. <https://doi.org/10.1007/s10654-014-9890-7>
- Hergenrath, K. C., Zeglin, R. J., McGuire-Kuletz, M., & Rhodes, S. D. (2015). Employment as a Social Determinant of Health: A Review of Longitudinal Studies Exploring the Relationship Between Employment Status and Mental Health. *Rehabilitation Research, Policy, and Education*, *29*(3), 261–290. <https://doi.org/10.1891/2168-6653.29.3.261>
- Janssen, B. G., Madhoum, N., Gyselaers, W., Bijmens, E., Clemente, D. B., Cox, B., Hogervorst, J., Luyten, L., Martens, D. S., Peusens, M., Plusquin, M., Provost, E. B., Roels, H. A., Saenen, N. D., Tsamou, M., Vriens, A., Winckelmans, E., Vrijens, K., & Nawrot, T. S. (2017). Cohort Profile: The ENVIRonmental influence ON early AGEing (ENVIR ON AGE): a birth cohort study. *International Journal of Epidemiology*, dyw269. <https://doi.org/10.1093/ije/dyw269>
- Krokstad, S., Langhammer, A., Hveem, K., Holmen, T., Midthjell, K., Stene, T., Bratberg, G., Heggland, J., & Holmen, J. (2013). Cohort Profile: The HUNT Study, Norway. *International Journal of Epidemiology*, *42*(4), 968–977. <https://doi.org/10.1093/ije/dys095>
- Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., Perola, M., Ng, P. C., Mägi, R., Milani, L., Fischer, K., & Metspalu, A. (2015). Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *International Journal of Epidemiology*, *44*(4), 1137–1147. <https://doi.org/10.1093/ije/dyt268>
- Mackenbach, J. P., Roskam, A.-J. R., Schaap, M. M., & Menvielle, G. (2008). Socioeconomic Inequalities in Health in 22 European Countries. *N Engl J Med*, *14*.
- Mimilidis, H., Demarest, S., Tafforeau, J., & Van Der Heyden, J. (2014). *Projet de couplage de données issues de l'enquête de Santé 2008 et des Organismes Assureurs*.
- Naess, O., Sogaard, A. J., Arnesen, E., Beckstrom, A. C., Bjertness, E., Engeland, A., Hjort, P. F., Holmen, J., Magnus, P., Njolstad, I., Tell, G. S., Vatten, L., Vollset, S. E., & Aamodt, G. (2008). Cohort Profile: Cohort of Norway (CONOR). *International Journal of Epidemiology*, *37*(3), 481–485. <https://doi.org/10.1093/ije/dym217>
- Nguyen, D., Hautekiet, P., Berete, F., Braekman, E., Charafeddine, R., Demarest, S., Drieskens, S., Gisle, L., Hermans, L., Tafforeau, J., & Van der Heyden, J. (2020). The Belgian health examination survey: objectives, design and methods. *Archives of Public Health*, *78*(1), 50. <https://doi.org/10.1186/s13690-020-00428-9>
- Renard, F., Devleeschauwer, B., Van Oyen, H., Gadeyne, S., & Deboosere, P. (2019). Evolution of educational inequalities in life and health expectancies at 25 years in Belgium between 2001 and 2011: a census-based study. *Archives of Public Health*, *77*(1), 6. <https://doi.org/10.1186/s13690-019-0330-8>

REFERENCES

- Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J., Dotinga, A., Vonk, J. M., van Dijk, F., van Zon, S. K., Wijmenga, C., Wolffenbuttel, B. H., & Stolk, R. P. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, *44*(4), 1172–1180. <https://doi.org/10.1093/ije/dyu229>
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
- Tsao, C. W., & Vasan, R. S. (2015). Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *International Journal of Epidemiology*, *44*(6), 1800–1813. <https://doi.org/10.1093/ije/dyv337>
- Van Beckhoven, D., Buvé, A., Ruelle, J., Seyler, A., Sasse, A., & Belgian HIV Cohort Study Group. (2012). *A national cohort of HIV-infected patients in Belgium: design and main characteristics*. *67*(5), 333–337.
- Van den Eeden, L., Lambrechts, N., Verheyen, V., Berth, M., Schoeters, G., & Jacquemyn, Y. (2018). Impact of particulate matter on mothers and babies in Antwerp (IPANEMA): a prospective cohort study on the impact of pollutants and particulate matter in pregnancy. *BMJ Open*, *8*(3), e020028. <https://doi.org/10.1136/bmjopen-2017-020028>
- Van den Eynden, J., Descamps, T., Delporte, E., Roosens, N. H. C., De Keersmaecker, S. C. J., De Wit, V., Vermeesch, J. R., Goetghebeur, E., Tafforeau, J., Demarest, S., Van den Bulcke, M., & Van Oyen, H. (2018). The genetic structure of the Belgian population. *Human Genomics*, *12*(1), 6. <https://doi.org/10.1186/s40246-018-0136-8>
- Van Oyen, H., Charafeddine, R., Deboosere, P., Cox, B., Lorant, V., Nusselder, W., & Demarest, S. (2011). Contribution of mortality and disability to the secular trend in health inequality at the turn of century in Belgium. *The European Journal of Public Health*, *21*(6), 781–787. <https://doi.org/10.1093/eurpub/ckq198>
- Van Oyen, H., Tafforeau, J., Hermans, H., Quataert, P., Schiettecatte, E., Lebrun, L., & Bellamammer, L. (1997). *The Belgian Health Interview Survey*. *55*, 1–13.
- Vartiainen, E. (2018). The North Karelia Project: Cardiovascular disease prevention in Finland. *Global Cardiology Science and Practice*, *2018*(2). <https://doi.org/10.21542/gcsp.2018.13>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Willemsen, G., Baselmans, B., Sinke, M., Bovenberg, J., Nieuwboer, R., Helmer, Q., Bernard, M., & Boomsma, D. (2014). *Individual feedback on personal research results via the Netherlands Twin Register Portal: Participants love it!* International Congress on Twin Studies.
- World Health Organization. (2020). *Novel coronavirus(2019-nCoV) Situation report* (No. 21). https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200210-sitrep-21-ncov.pdf?sfvrsn=947679ef_2.
- Zenger, M., Berth, H., Brähler, E., & Stöbel-Richter, Y. (2013). Health complaints and unemployment: The role of self-efficacy in a prospective cohort study. *Journal of Social and Clinical Psychology*, *32*(1), 97–115. <https://doi.org/10.1521/jscp.2013.32.1.97>

ANNEXES

1. Key meetings

1.1. STAKEHOLDERS

Agence wallonne pour une vie de qualité – AVIQ
Agentschap Zorg en Gezondheid
Belgian Science Policy Office – BELSPO
Commission communautaire française – COCOF
Deutschsprachige Gemeinschaft
Fédération Wallonie-Bruxelles
Federal Planning Bureau
Federaal Kenniscentrum/Centre fédéral d'expertise - KCE
Fonds de la Recherche Scientifique – FNRS
FPS Social Security
Gemeenschappelijke Gemeenschapscommissie van Brussel-Hoofdstad / Commission communautaire commune de Bruxelles-Capitale – GGC/COCOM
Institut Scientifique de Service Public – ISSeP
Institute Walloon de l'évaluation, de la Pprospective et de la statistique – Iweps
InterMutualistisch Agentschap/ L'Agence InterMutualiste – IMA/AIM
Katholieke Universiteit Leuven – KU Leuven
Observatoire de la Santé – Liège
Observatoire de la Santé et du Social Bruxelles
Observatoire de Santé de province de Namur
Observatoire de Santé du province de Hainaut
Province de Hainaut
Rijksinstituut voor ziekte- en invaliditeitsverzekering/Institut national d'assurance maladie-invalidité – RIZIV/INAMI
Statistics Belgium - StatBel
Statistiek Vlaanderen
Superior Health Council
University Antwerp – UA
Université catholique de Louvain – UCL
University Gent – Ugent
University Hasselt – UHasselt
Université Libre de Bruxelles – ULB
Université de Liège – ULg
Vlaams Instituut Gezond Leven
Vlaamse Instelling voor Technologisch Onderzoek – VITO
Vrije Universiteit Brussel – VUB

1.2. SUMMARY KEY MEETINGS

Exploratory meetings (2016/2018)

Aim and focus

- It is important to determine what the goal of this cohort should be. This will increase support for such an endeavor.
- The focus of the cohort should be on the current public health challenges: The ageing of the population is a good example of this (healthy ageing), as well as health care seeking behavior, mental health, 'omics', migration, inclusion of *new people* (migrants) (this could be a selling point and a novelty compared to other existing cohorts), well-being: happiness/life course events/sense of belonging.
- However, the focus should not be too broad or too much 'multipurpose': identify 'hot topics' and objectives that target the major challenges/hot topics in Belgium: families or healthy ageing for instance. One selling point should be defined. At first glance, the topic of healthy ageing is very interesting, but should not be competing with data that is already available, for example in the SHARE-project.
- Both research and policy will have to be involved when it comes to determining the focus of the cohort.
- The question is raised whether this instrument should be able to be of use for policy evaluation. If such a cohort can be used to measure the impact of policy-based interventions, this is a great added value.
- The Precision Medicine Initiative of the National Institutes of Health in the US provides a nice example: the focus is on longitudinal studies in both sick and healthy individuals, in order to map the 'path to disease'.
- Health care and health policy have only about 10% impact on health status. From the examples in cohorts broad it appears that much attention is paid to (and much money invested in) biomarkers, MRI, genetic markers, etc. but data on major drivers of population health such as housing, labour market, income redistribution,.. are also essential. If information on social conditions is included in a population cohort there will be a broader support.
- In order to guarantee the sustainability of the cohort, a short term focus should be formulated (which can be used as selling point), as well as a long term focus.

Balance novelty / complementarity

- In an optimal research design, data collection should, on one hand, be comparable to cohorts abroad and harmonization of the data should be possible in order to make comparisons and evaluate national public health policies, however, on the other hand, it should have novel components, to make the cohort unique.
- Collecting the same information in a Belgium cohort as in cohorts abroad might strike as counterintuitive, but can aid in mapping the differences in health (policies) in European countries and can support the cost-efficiency of the cohort, as data can be used by researchers abroad. The use of standardized methods can be justified by the need of comparability and increase of statistical power.
- Cohort data should not only be compatible with cohort data abroad, but also with the data collected within the cohort itself: Baseline levels of, for example, lab readings should be comparable to lab readings collected in future waves.
- It is of importance that a cohort and existing initiatives, should be complementary, rather than competitive. The focus should be on a new set of questions that needs to be answered. For example, there is already an experience with a population cohort through the interface of Demography of VUB, which does research activities based on a mortality follow up of the population census 1991, 2001 and 2011. The database used by the Interface Demography is managed by Statistics Belgium. Statistics Belgium is convinced that further steps are required

to make research on those data possible, integrating also data from other databases. In addition, the Intego database managed by the Academic Centre for General Practice of the KULeuven has data on 450.00 patients. There are plans to make linkages with health insurance data and environmental health data. This is realized with quite limited budgets. It would be good to optimize those type of existing databases, rather than starting from scratch.

Administrative data

- As there are many sources of administrative data available in Belgium, setting up a retrospective cohort could be an option.
- Due to the large amount of administrative data available, choices have to be made. The 'Belcohort' could take the form of a 'production line': linking data from different sources, thereby following a large group of individuals in an administrative way. Results from smaller studies, such as the HIS, can be linked to this and smaller or nested studies can be performed on this production line as satellites.
- In addition to data from the mutualities/register of births/StatBel etc., data can also be obtained from electronic records of hospitals/general practitioners and other healthcare providers.
- As far as the costs are concerned, an administrative cohort, in which the data is supplemented and enriched from other sources, will be more cost-efficient than setting up a cohort from scratch.

Structure of the cohort

- The cohort could take the shape as an open research infrastructure or consortium. By doing so, anyone can access the data to analyze health issues, support health guidelines or evaluate health policies. In other countries, the data collected in these population-based cohorts is also available to (international) researchers.
- The minimum size of such population cohorts could be between 50,000 and 100,000. If we take all the Health Surveys that have been carried out in Belgium over the past 20 years together, we have already reached this number. A population cohort should also offer opportunities for implementation research projects, e.g. nested studies to test new types of care.
- We will have to think carefully about the governance structure: not only administrative, but also technical issues should be countered.
- New evolutions, such as Internet panels, need to be considered. These are cohorts, mostly of a few thousands correspondents, which cover also non-health related topics. Participants receive a laptop and broadband access, and receive questionnaires about small modules. However, at this stage the relevance of Internet panels in the framework of the exercise that is planned is a detail.
- Cohorts abroad differ in ways of recruitment: some collect data within families, others cohorts only comprise of non-related individuals. The pros and cons should be considered carefully.
- Many population based cohorts collect data in waves: the average time between the waves is 5 years, but this might be too long. Smaller, in-between-waves (web-based) questionnaires could solve this problem
- When adherence is low and drop-out of participants is high, new groups of participants should be included (refreshment sample).
- The cohort should be presented as a research infrastructure, which is easily accessible. The ESS might provide a good example as it consist of fixed modules as well as 'rotating' modules. This flexibility is a worthy selling point: not only can it provide the opportunity to answer the current *burning* questions, it might also appeal to future policy makers as they can also reap the benefits from this infrastructure by adding new modules.

Added value

- Other countries have successfully set-up population-based cohorts, this indicates that there is a need for infrastructures like these cohorts. However, as there are differences between

Western-European countries, data collected within these cohorts abroad might not be generalizable to Belgium. A large cohort in Belgium might be key in supporting as well as evaluating Belgium policies, not only on the federal level, but also on the level of the different regions in Belgium. Moreover, if there are no Belgian-specific data, we risk not to be involved in international projects and have no impact on the research agenda or developments in health care.

- The HIS (health interview survey, cross-sectional population survey, in-person interviews) lacks the longitudinal aspect. In addition, the HIS lacks the medical examinations that are currently taking place within cohorts abroad. Research could benefit from a cohort regarding its longitudinal aspect: for example, research into substance abuse (in minors) and risk factors for substance abuse later in life or consequences of substance abuse can be mapped with precision. Moreover, the strength of combining a longitudinal design with a large sample size is that causality can be explored, taking into account a wide range of determinants in many fields : genomics, lifestyle, social and physical environment, ...
- As a cohort can map longitudinally exposure-outcome relationships, and thereby mark risk factors that might be amendable for intervention, it can be cost-efficient. Although it is difficult to provide an estimation, establishing such a population-based cohort might be more cost-effective in the long run (compared to a scenario in which Belgian cohort-data is not available).
- The administrative data that is already available lacks qualitative data: for example, with regard to research into unemployment, information is missing on type of work/employment, which can be linked to lifestyle factors (focusing on employment). In international databases this information is lacking as well. Information on the relationship between poverty and health or migration and diseases in a longitudinal design would be very welcome.
- One population-based cohort might reduce the costs of maintaining small local initiatives, as these can be incorporated in the infrastructure.

Accessibility

- It is argued that both for existing and new data sources more support will be found if there is a push policy which promotes data accessibility. Data access should be easy and for free. Data should be available one year after data collection. Organizations should be evaluated based on the use of the data. This is not only the case for health data. Data access in Belgium is a problem. It is often much easier to obtain data from other countries than Belgian data.

Engagement

- In order to increase response rate, the cohort should invest in engagement and sensibility of the participants.
- Initial response rate in other cohorts, such as CONSTANCES in France is very low. The biggest challenge regarding response rates might not lie in the follow ups, but in the first wave of data collection.
- In addition, to keep the 'cohort alive' time and money should be invested in keeping the attention of the participants, for example by sending newsletters, Christmas cards etc. and the use of social media.
- It is worthwhile to seek for an engagement for 2 or 3 waves from the beginning.
- will be necessary to bring research groups together to arrive at specific questions. The involvement of citizen scientists and crowdsourcing is also important to take into account. The success of a cohort will depend on the exploitation of these issues.

International

- It is important that such a cohort aims at complementing already existing health information available within the EU. Data harmonization will therefore need attention. Data linkage is of importance in order to support and assess new economic solutions in the field of health.
- A number of successes have been achieved in linking data between European countries, but much remains to be done.

- European cooperation between national population cohorts, as it has grown for Health (EHIS) and Research (EHES) surveys, is not yet on the agenda, but will certainly be an option in the future.

Work group meetings (2019)

Added value

- What is needed is a *longitudinal* data collection effort
- The largest knowledge gaps could be solved with *linking data*; linking data that is currently in different 'silos'; for example, IMA/AIM data is missing diagnosis, only proxies are used. Linking data can aid in this.
- Data linkage should be in a *structural* manner. The current process of project-based linkage anymore costs a lot of time and resources.
- Surveys and clinical examinations should be incorporate in the cohort. Data linkage only is not sufficient. For example, information on *lifestyle, family and household characteristics, health care needs, life course events, labor conditions, child poverty, and (subclinical) mental health issues* is currently missing.
- Genomics become more important, *biobanking* should be an important focus point.

Starting point

- The Belgian Health Interview survey (BHIS) could be taken as starting point. Currently, the BHIS is being linked to IMA/AIM data. A 'belcohort' should be more ambitious than just linking data. Nevertheless, to start with a large sample of which lot of data is already available is favorable.
- The added value of the Health Examination Survey (HES) is that people were visited at home, this ensures a more representative sample; the HES could also serve as starting point or pilot study.
- Linking the larger databases such as IMA/AIM, hospital data, mortality data (StatBel), could be an interesting exercise; however, these data should be complemented by GP data (for example the Intego network), data on environment (for example air quality). Linking data is already technically feasible; the legal feasibility is more important.
- A platform could be created were all smaller initiatives and cohorts can be listed in a catalogue, as well as metadata on administrative data sources; this might not represent a true cohort, but will be an open research infrastructure. This platform could be extended towards a distributed research infrastructure, where data is not centralized, but pipelines for data linkage are build.
- An online pane might be a good and low costs alternative and may serve as a starting point. Public services and scientists can use this infrastructure to distribute surveys. It is a flexible tool that can run for a long time.

Privacy

- A legal framework will have to be created and application procedures for data requests should be set up.
- For only linking administrative data no consent is needed; this puts no limit on the sample size.
- When including local data sources such as GP data, the data providers might be hesitant to participate as individual data can be traced back to their practice.

Adherence & drop-out

- To avoid loss to follow up, we can learn from initiatives abroad; especially the communication with participants seem to be important, rather than (monetary) incentives (although generally a travel reimbursement is offered). Especially the feedback to participants is appreciated.
- Especially in Belgium the engagement of the participants will be difficult for a national initiative, as there are large regional differences.
- We should avoid subsamples being 'over interviewed', as this might lead to drop out.

Sample size & characteristics

- The sample size should be big enough in order to conduct research on subsamples (nested studies).
- Representativeness very important; also minorities should be included in the sample. Especially regions where the healthcare need is high are generally underrepresented in many studies.
- If drop out becomes high, we could include a refreshment sample.
- Sampling could be done in households, to address issues such as family and household characteristics, health care needs, life course events and child poverty. Many decisions are based in the context of a household; retirement, informal care, etc. There is a health effect of family environment.
- Although it might be interesting to start with a birth cohort, this might be too long to generate results. If they are part of the sampled household, they can be included in the data collection.

Aim of the cohort

- Generating data and/or making secondary data available.
- Evaluation of health policy; Policy wise the time lag is important; this should not be too long. Sometimes it takes years for data to be available.
- Standardizing indices, measures and variables of current data collection efforts. Also; validating collected data.
- Addressing urgent questions on 'hot issues', for example the causes of disability and unemployment.

Accessibility

- Data is only of interest to the targeted users if it is free and available. For example in the SHARE-study: if you apply for data, you have the data within a day.
- Access to data can be done in different ways; the Norwegian system is interesting as there is access to data but data stays on the local server of the data holder.

2. Workshop

International Workshop on cohort studies: to exchange experiences, know-how, ideas and expertise on the set-up and maintenance of population-based cohorts - 13 September 2019

2.1. AGENDA

09:00 – 09:30 Registration

09:30 – 09:40 Welcome by Herman Van Oyen

Part I: Cohort studies and research infrastructures in European countries

09:40 – 10:20 Lifelines (The Netherlands)

10:20 – 11:00 Swiss National Cohort (Switzerland)

11:00 – 11:30 Break

11:30 – 12:10 Microdata.no (Norway)

12:10 – 12:50 German National Cohort (Germany)

13:00 – 14:00 Lunch

Part II: Opportunities for public health research in Belgium

14:00 – 14:45 Introducing the panel

14:45 – 15:40 Panel discussion Belgian stakeholders

15:40 – 15:45 Wrap up by Nienke Schutte

Part III: Reception

15:45 – 17:30 Network event

2.2. SESSIONS

Lifelines (The Netherlands)

Lifelines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviors of more than 167.000 persons living in the North of The Netherlands. The Lifelines Cohort Study provides through its length, broad-ness and in-depth examination excellent opportunities for studies worldwide unraveling the etiology of multifactorial diseases focusing on multifactor risk factors.

Speaker: Ilse Broeders

Swiss National Cohort (Switzerland)

The Swiss National Cohort (SNC) is a long-term, population based multipurpose cohort and research platform. The current version of the SNC is based on census data from 1990 and 2000 that were linked to mortality, life birth and emigration records until 2015, and to the newly introduced Registry Based Census (RBC) and annual structural surveys from 2010 onward. The SNC enables research in a wide range of public health subjects, in particular in combination with other longitudinal or environmental data.

Speaker: Marcel Zwahlen

Microdata.no (Norway)

Norway has a well-developed and well-documented system for register data. Some of the most important registers are also used as a basis for official statistics. Such data is also valuable as research data. Microdata.no is operated and being developed by the Norwegian Centre for Research Data and Statistics Norway, and provides researchers and students at approved research institutions with access to register data from Statistics Norway.

Speakers: Johan Sjøberg & Trond Pedersen

German National Cohort (Germany)

The German National Cohort (GNC) has been inviting men and women aged between 20 and 69 to 18 study centers throughout Germany since 2014. The participants are medically examined and questioned about their living conditions. The GNC's aim is to investigate the causes of chronic diseases, such as cancer, diabetes, cardiovascular diseases, rheumatism, infectious diseases, and dementia in order to improve prevention, early diagnoses and treatment of these very widely spread diseases.

Speaker: Halina Greiser

A panel discussion

with policymakers, administrators and researchers to discover how a large population-based cohort study can play a key role in improving the health of the Belgian population.

Moderators:

Dimitri Mortelmans (University Antwerp)

Jean Tafforeau (Sciensano)

Panel:

Rudi Van Dam FPS Social Security

Dominique Dubourg AVIQ

Patrick Lusyne StatBel

Christoph Van Roelen VUB

Karin De Ridder Sciensano

2.3. CONCLUSIONS (SUMMARY)

- In Belgium there is a need for a cohort or similar research infrastructure, containing multiple variables, collected within a longitudinal design. The successes abroad show us that this need is not only specific for Belgium; on other West European countries larger cohorts have been set up. Despite their differences in recruitment, measurements, sample size etc., the common goal is to create a longitudinal data collection effort and making these data available for both research and policy. In order to address these target users, the societal relevance on the short term and the long term should both be kept in mind when designing and setting up the study protocol. A large sample size can allow for nested studies.

- The participants within the cohort should get an active role; we should aim to create a platform where the participant itself is the key person.
- There are many knowledge gaps that could be addressed in a cohort study; not only research gaps, but also policy gaps; about active ageing, poverty & well-being. Current funding initiatives are focused on diseases, but a cohort might focus more on well-being and social participation; important to focus on what actions are needed to prevent individuals from getting diseased or disabled.
- Current data requests for health data that are governed by other parties are regarded as 'a hassle'; data should be easily accessible. This can be established by designing a clear data application process and a platform where all metadata or aggregated data is listed. Both require a clear governance plan.
- As Belgium has many administrative data sources, the feasibility of linking these data sources in a structural manner should be explored. Data should be integrated, to come to an integrated health information system. This allows visualization of these data and information the current state of the Belgian health information system. Finally, this can aid in maximizing the use of administrative data sources and finally valorizing these existing data.
- Nevertheless, linkage initiatives are welcomed, but survey data containing information in subjective experiences are also very much needed. It is important not to focus all time and resources on the linkage aspect, as one can lose sight of the importance of clinical measures and survey data.
- Finally, there should be long term engagement and investments. Addressing both the urgent policy questions as well as a focus on modeling longitudinal causal relationships could help in 'selling' this initiative to funders

CONTACT

Karin De Ridder • T+32 2 >xxx xx xx • >first name.last name@sciensano.be

MORE INFORMATION

—

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

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

Responsible editor: Myriam Sneyers, Managing director a.i. • Rue Juliette Wytsmanstraat 14 • Brussels • Belgium • >D/xxxx/xxxx/xx