



FORESIGHT AND MODELLING FOR EUROPEAN HEALTH POLICY AND REGULATION

Project Acronym:	FRESHER
Project Full Title:	Foresight and Modelling for European Health Policy and regulation
Grant Agreement:	643576
Project Duration:	36 months (January 2015 - December 2017)
Coordinator:	AMU - Jean Paul Moatti

D 5.1 Validated European Health Policy Model software and documentation

Title:	Validated European Health Policy Model software and documentation							
WP	WP5 Modelling and policy simulation							
Authors:	Christoph AUBRECHT, Carla BLÁZQUEZ, David CANTARERO, Bruno CASAL,							
	Sebastien CORTAREDONA, Marion DEVAUX, Mario KÖSTL, Santiago LAGO,							
	Alienor LEROUGE, Audrey MICHEL-LEPAGE, Jean-Paul MOATTI, Alain							
	PARAPONARIS, Marta PASCUAL, Thierry PELLEGRINI, Francisco REYES,							
	Berta RIVERA, Franco SASSI, Klaus STEINNOCHER, Bruno VENTELOU							
Organisations:	AIT, AMU, OECD, GEN							
Dissemination:	Public							
Date of publication:18/07/2016								





Table of content

Model Description
Modelling Principles
General Architecture ϵ
Data Component6
MicroSimulation Framework (MSF)7
Demographic Component
Diseases and Risk Factors Component 9
Model Validation
Validation Principles
Demographic Module Validation11
Population Validation
Disease Module Validation
Population Validation
Prevalence Validation15
Health expenditures calculations 17
A - Disease-oriented approach
B - Socioeconomic determinants of non-communicable diseases21
Geospatial exposure analysis and city-level modeling25





Model Description

The FRESHER microsimulation model is designed to model the impacts of behavioral and metabolic risk factors on chronic diseases and longevity, as well as the extent to which specific policies can modify those impacts.

The model uses case-based microsimulation to create representative synthetic life histories from birth to death providing multiple cross-sectional representations of a population during the 'validation period' – when both simulated and historical data are available and can be compared to validate or calibrate the model (1990-2010) – and the 'projection period' – over which quantitative estimates of the future global burden of chronic non-communicable diseases (NCDs) in the EU and policy impact (2011-2030 and 2011-2050) are simulated.

The Microsimulation Framework is the C++ engine which creates and simulates the individuals. This global architecture also allows us to take into account the different mega trends and scenarios generated by the work of WP4.

Modelling Principles

The simulation creates a large synthetic population representative of a specific country or geographical zone, aggregation of individuals at a time, followed from birth to death. Birthdate, gender and immigration status and year of arrival are first initialized to take into account the distribution of the population. The individual is then simulated from birth to death and its life trajectory is marked by different "life events" such as emigration, death or disease incidence.

The model simulates life histories in continuous time. Each event can happen at any time (unlike in discrete-time modeling where events can only happen within given time intervals). In most cases, time to event is stochastically determined, based on an exponential distribution the intensity of which (λ) varies for different events and is a function of individual characteristics. When the intensity of the distribution changes, time-to-event is stochastically re-determined by the model based on the new parameters of the distribution. Events compete with each other, i.e. the shortest time to event will determine what event happens first. An event can modify individual characteristics and consequently impact the likelihood of other events occurring (by modifying the intensity of their distributions).







Figure 1: Time to event distribution, exponential distribution





Model Input

Input Database: Collection of data from different sources

- Demographic Data (HMD + UN)
- Epidemiological Data (IHME + IARC)
- Risk Factor Epidemiology Data (NCD's Risk, etc..)
- Etc...

Configuration

Years of validation and projection period, diseases List, risk factors list, scenario parameters, etc..

Mega Trends Parameters

Data Manager (py)

Simulation database Interface

Input Tables

- Demographic Module: Mortality Table, Births Table, Immigration Table
- Health Module: Incidence, Fatality, Remission, and Prevalence Tables.
- MultiRisk Factor Module: Risk Factor
- Prevalence Tables, Relative Risk Tables Etc..

Simulation Configuration

- Seed
- Number of cases
- Number of subsamples
- etc

MicroSimulation Framework (c++)

Simulation Component

Aim: Initializing and simulating all individual characteristics along time (implementation of characteristic and events)

Result Component

Create Result Tables from simulated trajectories. Tables are created during the simulation.

Final Outputs

Relevant Tables eg.

- Population Table
- Validation Tables
- Epidemiology outputs
- Health Expenditures
- Etc..

Diagram1: General Architecture of the International Alcohol Policy Model





General Architecture

Each component of Diagram1 is described in this section.

Data Component

Input database:

The **Input Database** is defined as the collection of different databases used by the model. As long as it is possible, the model relies on international datasets, which provide consistent estimates of parameters across countries. The following are some of the main data sources used:

- Human Mortality Database (HMD), publicly available at <u>http://www.mortality.org</u>.
- UN World Population Database (UNPD), publicly available at http://www.esa.un.org
- IHME Epidemiological Estimates, visualizations of the database are available at http://vizhub.healthdata.org/epi/

Estimates of relative risks derived from the work of WP2 are also stored in the input database.

Data Manager

The Data Manager is a tool implemented in Python. Its main functionalities are described below.

Data Extraction & Data Transformation

The data manager is responsible for extracting and transforming, when needed, the data for a specific country. The main output tables are the following:

- *Population Table*: extraction of HMD population records for the historical period plus extraction and scaling of population projections from UNPD for the projection period.
- *Mortality Rates Table*: extraction of HMD central dates rates for the historical period. For the projection period the model uses the Life Expectancy Projection from UNPD to extrapolate the last historical data available.
- *Births Records Table*: extraction of HMD birth records for the historical period. For the projection period, the computed *Population Table* and Crude Birth Rate projection from UNPD are used to compute the Births Projections.
- Immigration Table: For each year of the estimation and projection periods, the model calculates the net migration component by age and gender from population records as the population variation that remains unexplained after accounting for births and deaths. Positive components are seen as inward migration whereas negative components account for outward migrations, based on the following equation: P(n + 1, y + 1) = P(n, t) D(n, t) + I(n, t), with P(n, y), the population of age n at year y, D(n, y) the number of deaths, and I(n, t) the net immigration.
- *Epidemiological Tables*: Incidence, Excess Mortality and Remission estimates are interpolated from the IHME dataset and integrated to compute the Epidemiological Tables.

Data Aggregation

The data are extracted and transformed at the country level and then the countries are aggregated into the three European regions used in the FRESHER project. The aggregation is built on the following principles.

Mega Trend Scenario Builder

The data manager is also responsible to use the Mega-trends Scenario to build different projections according to the evidences found by WP4.





Simulation Database Interface

The output of the Data Manager is a sql database which contains all of the "transformed and aggregated data" for a country/geographical zone, plus all the parameters the engine needs to run: number of cases, time intervals for both validation and projection period, etc.

MicroSimulation Framework (MSF)

MSF is the core of the microsimulation process. Its implementation allows us to simulate as many virtual individuals as needed. It also allowsto produce various outputs, as rates of disease prevalence or survival rates, based on a configuration defined by the user and stored in the Simulation Database Interface. The architecture has been done such as the initialization of an individual is independent of the events. The main c++ classes implemented in MSF are briefly presented below.

Main C++ class

class msf::DiscreteState

It's the base class for the implementation of all the individual characteristics. Its main derived classes are the following:

- class msf::Age
- class msf::Resident records the resident status
- class msf::Gender
- class msf::Alive
- **class** msf::Illness records the disease status
- class msf::Obesity records the BMI Category (Normal Weight, Over-Weight, Obese)
- class msf::SmokingStatus records the Smoker Status (Smoker, Non Smoker)
- class msf::DeathCause records the cause of death

class msf::Actor

It's the class which represent the individual. Basically it contains an instance of each DiscreteState which describes the individual characteristics.

class msf::SimulationEvent

It's the base class for the implementation of all the events. The main attribute of this class is the date when the event occurs. Its main method is ProcessEvent(EventQueueController& controller) which implements the events and its consequences on the individual characteristics -e.g. the event impacts Illness status of the individual. One of its members is a reference on an Actor, the event can be notified when one of the **DiscreteState** of the Actor has changed. Its main derived classes are the following:

- class msf::BirthdayEvent
- class msf::DeathEvent
- class msf::EmigrationEvent
- (depending on the Illness status).
- implements the death by "other causes"
- implements the emigration event. implements disease incidence and disease remission
- class msf::DiseaseToggleEvent
- class msf::DeathByDiseaseEvent implements the disease fatality

class msf::EventQueue<Event>

It's a container of SimulationEvent instances which is responsible for the competing event framework by keeping those instances ordered by their time to event and processing the next event.





class msf::ActorFactory

It's the class which is responsible to initialize an instance of Actor (the individual) and its associated EventQueue<Event>.

class msf::Engine

It's the class which is responsible to run the simulation. For every cases of the simulation, a new instance of Actor is instantiate through the ActorFactory and the EventQueue is processed until one of the exit event happened (death or outward migration).

Demographic Component

Initialization of individual demographic characteristics: birthdate, gender, immigrant status and date of arrival.

The model is designed to produce - by simulation - a synthetic representation of a population over a range of years instead of starting from a cross sectional population at a given point in time. Individual characteristics are then not inputted at a certain date but simulated through the model. They can therefore be "observed" as they are generated by the model, and their simulated distributions can be compared with historical or projected data.

The key point of this microsimulation exercise is to parametrize the initialization of an individual at birth and his/her various life events, in order to have a synthetic population which is statically coherent with input data. The birthdate, the gender and the inward migration are treated as an individual characteristics assigned at birth.

Example: Initialization of individual characteristic: Birthdate(in case of no immigration)

- Parameter Range : As people are simulated from birth to death, in order to have all ages (0 to 110) represented from 1900 to 2050, **Birthdate** range will be: [1890, 2050]
- Input Distribution: Birth records (historical and projected) are used to create the birth year distribution of cross-sectional populations.
- Individual assignment: A Birth Year is randomly assigned to an individual according to the Birth Year distribution. A random noise is then added to this birth year to get the Birthdate.

From the inward migration pattern data, the proportion of immigrants in the total population is derived, along with the distribution of their individual characteristics (gender, birthdate and year of arrival).

Mortality

All individuals in the model have a death event, whose timing depends on a mortality rate. For the historical period, this rate is directly derived from the Human Mortality database which provides average death rates by age and gender for a large historical period. For the projection period, the last historical data are extrapolated to fit UN life expectancy projections.

The death rate is modified to take into account the fatality of all the diseases included in the simulation.



Outward migration

All individuals in the model who are currently resident in a country have a given probability of leaving the country as a result of outward migration, which is -date-, age- and gender-specific.

Since detailed information on the direction of migration flows between countries in the same "region" is not available, country data on migration flows are not used in order to avoid double counting migrants.

Diseases and Risk Factors Component

Incidence, Remission and Fatality Event

For each disease there is a disease incidence event. The hazard rate is defined in the epidemiological tables of the Simulation Database Interface. Rates are age, gender and year specific. When the disease incidence event happened, it allows the fatality event and the remission event to compete.

The incidence hazard rate can be impacted by some of the Risk Factor of the individual. For each risk factor group, the conditional incidence is computed.

Example: IHD incidence for smokers and non-smokers

The parameter $\mathbb{P}(IHD)$ is the total incidence of IHD, $RR_{S/\bar{S}}$ is the relative risk of IDH for smokers compared with non-smokers and $\mathbb{P}(S)$ is the prevalence of smoking. The following formula is then derived:

$$\mathbb{P}(IHD|S) = \frac{RR_{S/\bar{S}}\mathbb{P}(IHD)}{1 + \mathbb{P}(S)(RR_{S/\bar{S}} - 1)}$$

It is assumed that those conditional incidences do not change over time. The total incidence changes are then due to changes in risk factors.

Risk Factors Modelling: BMI example

BMI is modelled as a continuous variable. $BMI = C_1 + C_2U$, where U, follows a continuous distribution (Beta or Log-Normal) C_1 , C_2 and the parameters of U are calibrated to fit some of the key parameters of the BMI distribution – eg the mean, the percentage of overweight and obese.

At birth, any individual is assigned a quantile which determines its position in the cross-sectional distribution by age and gender. It is assumed that this quantile will not change during the life of the individual. The evolution of BMI is then only due to ageing.







In Figure 3, for example, the individual positioned at the 60th percentile will have a BMI of around 26 between age 55 and 59 and around 28 between age 60-65.

Most of the risk factors are modeled with this framework which allows us to have simulated cross-sectional distribution coherent with input data.





Model Validation

Validation Principles

The model creates representative synthetic life histories from birth to death providing a cross-sectional representation of the population during the validation and the projection period. During the validation period both the simulated population and the historical one are available and can be compared to assess the ability of the model to coherently reproduce the cross-sectional distribution of the population.

The model is being validated at each step of the modeling process. First, the demographic component is validated, then the diseases modules and the risk factors. A table used for the demographic component will always be used for the validation of the other modules to be sure that changing model doesn't affect the quality of the model.

For this report, the analysis has focused on three countries, one per 'region': Germany for the Northern Region, Estonia for the Central-Eastern Region and France for the Southern region. The simulations have been run with 16 million of individuals.

Demographic Module Validation

Population matching: average error by age

In order to validate the Demographic module the differences between the actual population as observed in the data and the simulated population are examined. To evaluate those differences, the average relative error by age has been plotted throughout the validation period, as defined by the following formula:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{P}(n,y_i) - P(n,y_i)}{P(n,y_i)}$$

Where $\hat{P}(n, y_i)$, $P(n, y_i)$, is the simulated, the historical, population (age is n) at year $y_i \in [1990 - 2010]$.

The above graphs (Figure 3, Figure 4 and Figure 5) are plotting the relative error for, respectively, Estonia, France and Germany. The difference between the historical population and what is simulated through the model is very small, around 0.1% for most ages. The gap is higher for older ages, which is expected, as the population records become smaller (increasing the noise).





Population Validation



Figure 3: Demographic Module Validation, relative error on population records by age for Estonia

Figure 4: Demographic Module Validation, relative error on population records by age for France

Figure 5: Demographic Module Validation, relative error on population records by age for Germany

Disease Module Validation

Once the Demographic Module is validated, the disease module is introduced. Individuals can then develop a disease and then pass through the fatality or remission event. To validate this component, an assessment is made of how the disease module affects the previous results (on the population records) and how the model is able to reproduce the disease history by looking at the prevalence of the disease. The disease trajectory of an individual is determined by the incidence, the fatality and the remission of a disease as expressed in the data. Prevalence is then an output of the model.

Population Validation

The first step of the validation is to look at the impact of the disease module on the relative error between the historical population and the simulated as defined previously. The same measure as in Figure 3, Figure 4 and Figure 5, is used, when the disease module is activated and without.

Figure 6, Figure 7Figure 8 show that the relative error increases but remains small, leading to the conclusion that disease fatality has been accurately modelled.

Figure 6: Disease Module Validation, relative error on population records by age for Estonia with and without the disease module.

Figure 7: Disease Module Validation, relative error on population records, by age for Germany with and without the disease module.

Figure 8: Disease Module Validation, relative error on population records by age for France with and without the disease module.

Prevalence Validation

As explained previously, prevalence is an output of the model that can be used to validate the model. In the above graphs (Figures Figure 9Figure 10 and Figure 11) the prevalence of Ischemic Heart Disease (IHD) by age for men and women has been plotted. Dotted curves are the data as observed in the IHME data, solid curves are the output of the model. The two curves are very close, although a small degree of noise can be observed.

Figure 9: IHD Prevalence (simulated and historical) for Estonia in 2005

Figure 10: IHD Prevalence (simulated and historical) for France in 2005

Figure 11: IHD Prevalence (simulated and historical) for Germany in 2010

Health expenditures calculations

A - Disease-oriented approach

The main objective of this work is to estimate the cost of the chronic diseases that will be included in the microsimulation model developed within the FRESHER project. These cost estimations could be later used to calibrate the microsimulation model. France has been selected to implement and test a method which uses medico-administrative database. We started from (EGB) of the French National Health Insurance regime.

The EGB (Echantillon Généraliste des Bénéficiaire) is a permanent, representative sample of the population covered by National Health Insurance. It contains anonymous sociodemographic and medical characteristics and records of health care reimbursements. It was created using a systematic sampling method (1/97) on the two-digit control key of beneficiaries' national identification number. In 2014, the EGB was composed of 614,806 beneficiaries.

As a reminder, the chronic diseases to be included in the microsimulation model are heart disease, stroke, cancers (with a special focus on lung, colorectal, stomach and breast cancers), diabetes, COPD, cirrhosis and major depression.

In order to quantify the cost of each selected chronic diseases, a bottom-up approach is conducted for each chronic disease. In a bottom-up design, units of health care are used on a patient level and are multiplied with a price for this unit (1). All individual costs are then summed up to calculate total cost of the disease. Compared to a top-down approach, in which total expenditure for a given area or policy are divided by total units of activity, the bottom-up approach provides a greater level of accuracy.

However, in the French health care system, healthcare expenditures cannot be directly associated with a specific disease making a standard bottom-up approach infeasible. To overcome this limitation, we choose to estimate the cost associated with each chronic disease using regression models. Costs are estimated as the mean marginal difference of the predicted outcome with a chronic disease dummy switched on or off (see below). This allows for estimating the 'counterfactual' of what the cost would have been in the absence of the chronic disease while leaving the other model parameters unchanged. This approach is commonly used to estimate incremental costs for select diseases and risk factors (2-4).

Within each strata (=combinations of the covariates), the cost due to the chronic disease will be estimated as the mean marginal difference of the predicted outcome with the chronic disease dummy switched on or off :

$$c_{id_j} = c_{i|d_j=1} - c_{i|d_j=0}$$

with:

 $c_i d_j$ = cost associated with chronic disease d_j in strata i

 $c_i/d_j=0$ = predicted cost of hospital and ambulatory care in strata i for individuals without disease d_j

 $c_i/d_i=1$ = estimated cost of hospital and ambulatory care in strata i for individuals with disease d_i

One of the main advantage of the methods is that it is possible to take into account the comorbidity issue. For each chronic disease, the average cost per capita of the chronic disease will be estimated for people without any comorbidity and for people with. You will find below the first results, obtained on the EGB sample.

Costs per capita associated with each chronic disease

Among individuals with no comorbidity (among the selected chronic diseases – table 14):

- Except for diabetes, the costs associated with each chronic disease is the highest for incident cases (=diagnosis in 2014). For cancer for example:
 - Cost estimated at 11 002 € the year of diagnosis (2014)
 - Cost estimated at 7 772 € the second year after diagnosis (2013)
 - Cost estimated at 3 112€ if diagnosis before 2013
- Lowest costs observed for COPD and diabetes (< 2000€)
- Highest cost observed for cancers (> 10 000 €)

Average cost per capita and 95 % CI associated with the chronic diseases among persons with no comorbidities (2014 data - n = 481061)

	Age of disease < 1 year (diagnosis in 2014)			Age of disease = 1 year (diagnosis in 2013)			Age of disease> 1 year (diagnosis < 2013)		
	Mean	95 % confidence interval		Mean	95 % coi inte	nfidence rval	Mean	95 % coi inte	nfidence rval
COPD	1 740 €	1 733 €	1 747 €	880 €	875€	886€	997€	994€	1 000 €
Cancer	11 002 €	10 980€	11 025 €	7 722 €	7 703 €	7 741€	3 122 €	3 117 €	3 127 €
Cancer*	17 996 €	17 891€	18 101 €	13 638 €	13 516€	13 760€	4 459 €	4 437 €	4 480 €
Breast Cancer	11 265 €	11 213€	11 316€	9 086 €	9 040 €	9 132 €	2 638 €	2 625 €	2 652 €
Diabetes	1 389 €	1 379 €	1 399 €	1 152 €	1 145€	1 160 €	1775 €	1 772 €	1 778€
Heart disease	7 522 €	7 507 €	7 537 €	2 728 €	2 720 €	2 736 €	2 223 €	2 219€	2 227 €
Major depression	3 177 €	3 107 €	3 247 €	923€	919€	927€	1274 €	1 273€	1 276€
Stroke	6 965 €	6 916 €	7 014 €	2 401 €	2 378€	2 424 €	1917 €	1 901€	1 934 €

*Lung, colorectal and stomach.

Confidence intervals calculated using bootstrap with 1000 replications

Average cost per capita (€) and 95 % CI associated with the chronic diseases among persons with at least one comorbidity (2014 data – n = 481 061)

	Age of disease < 1 year			Age of disease = 1 year			Age of disease> 1 year		
	(diagnosis in 2014)			(diagnosis in 2013)			(diagnosis < 2013)		
	Mean	95 % confidence ו interval		Mean	95 % confidence interval		Mean	Леап 95 % confidence interval	
COPD	8 464 €	8 439 €	8 488 €	4 699 €	4 681€	4 718€	3 338 €	3 329 €	3 347 €
Cancer	15 918 €	15 865€	15 970 €	10 888 €	10 856€	10 921€	5 560 €	5 542 €	5 578€
Cancer*	23 253 €	23 138€	23 368€	19 508 €	19 265 €	19 751€	6 792 €	6 723 €	6 860 €
Breast Cancer	11 739 €	11 652 €	11 826€	8 816€	8 755 €	8 877 €	2 903 €	2 880 €	2 925 €
Diabetes	5 098 €	5 076€	5 121 €	3 400 €	3 373€	3 426 €	3 678 €	3 669 €	3 686 €
Heart disease	12 343 €	12 317€	12 369€	6 349 €	6 329€	6 369 €	5 082 €	5 072 €	5 093 €
Major depression	3 297 €	3 183 €	3 412 €	1 630 €	1 619€	1641€	2 231 €	2 225 €	2 237 €
Stroke	12 250 €	12 190€	12 310€	6 751 €	6 696 €	6 806 €	2 709 €	2 688 €	2 731 €

*Lung, colorectal and stomach.

Confidence intervals calculated using bootstrap with 1000 replications

Except for breast cancer, the costs associated with each chronic disease are much higher among individuals with at least one comorbidity. For diabetes for example:

- Cost estimated at 1 775€ when no comorbidity (date of diagnosis before 01.01.2013)
- Cost estimated at 3 678€ when at least one comorbidity (date of diagnosis before 01.01.2013)

References

1. Bound Co-CfSI. [cited; Available from: https://data.gov.uk/sib_knowledge_box/top-down-and-bottom-unit-cost-estimation

2. Bilger M, Finkelstein EA, Kruger E, Tate DF, Linnan LA. The effect of weight loss on health, productivity, and medical expenditures among overweight employees. Medical care. 2013 Jun;51(6):471-7.

3. Finkelstein EA, DiBonaventura M, Burgess SM, Hale BC. The costs of obesity in the workplace. Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine. 2010 Oct;52(10):971-6.

4. Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. Circulation. 2011 Mar 1;123(8):933-44.

B - Socioeconomic determinants of non-communicable diseases

If we focus on all the socioeconomic variables, the relationship between income and health is probably the most complicated (Fuchs, 2004). The correlation coefficient, obtained from the crudest associations, can range from highly positive to slightly negative, depending on the context and the aggregation level. Even when the positive correlation is strong and stable, causal interpretations may include income influencing health, health influencing income and/or "third variables" affecting both indicators in the same direction and at the same time. In addition, there is a large and growing body of literature in which the effects of income on health are examined because of the importance of these effects in the development of appropriate economic policies (Gravelle et al., 2002). Many studies findings suggest that individual health is a function of individual income – *the absolute income hypothesis*. In relation to income inequality, the *relative income–health hypothesis* suggests that income inequality has a detrimental effect on population health because it is an individual's relative, rather than absolute, income that is important for health (Eberstadt and Satel, 2004).

A systematic literature search was performed in PubMed, Cochrane Library and Web of Science (until 2015) to identify the most relevant published evidence regarding the relationship between income and health. In all databases, terms related to "health", "income" and "inequalities" were combined (for full search queries see Table 1). The searches were confined to papers published in the English language since 2010, to limit the scope of this review to the most recent data and the state of the art. In other words, we considered a 5-year retrospective horizon to be enough.

#	Search term						
PubMed							
#1.	Health [Title/Abstract]						
#2.	Income [Title/Abstract]						
#3.	Inequality [Title/Abstract]						
#4.	Limit to: journal article; year of publication >= 2010; English and Spanish; Humans subjects, free-full text.						
	Cochrane Library						
#1.	Health [Title/Abstract]						
#2.	Income [Title/Abstract]						
#3.	Inequality [Title/Abstract]						
#4	Limit to: year of publication >= 2010.						
	Web of Science						
#1.	Health [Topic]; [Title]						
#2.	Income [Topic]; [Title]						
#3.	Inequality [Topic]; [Title]						
#4.	Limit to: journal article; year of publication >= 2010; English and Spanish; Public Environmental						
	Occupational Health "or" Social Issues "or" Health Care Sciences Services.						

TABLE 1. Search strategy: PubMed, Cochrane Library and Web of Science

After finding publications in the electronic searches, duplicate records were removed. The selection of papers was ultimately based on the following eligibility criterion: an applied study with a focus on one or more OECD countries (included the European Union and other developed countries). Additionally, the results of "hand searching" are also included in the following pages, where a wider horizon is considered. Figure 1 is a diagram of the paper selection process following PRISMA (<u>www.prisma-statement.org</u>).

FIGURE 1

Flow diagram of paper selection process

The literature search located 291 publications in the databases under consideration, and 17 papers published between 2010 and 2015 were identified through "hand searching". A total of 11 duplicates were removed, resulting in 297 "unique papers". After screening the titles against the eligibility criteria, 90 papers were selected. Of these, 57 articles were excluded as they did not fit with the previous criteria. So, a final set of 33 selected studies have been taken into account in this review. In any case, further papers are finally considered to have a robust overview. The following Table 2 focuses on some recent selected studies (Vallejo-Torres et al., 2014; Torre and Myrskylä, 2014; Chauvel and Leist, 2015; Jutz, 2015; Lillard et al., 2015; Rambotti, 2015) extracted from the 22 papers found through the database search.

TABLE 2 Characteristics of some recent selected studies included in the review (omitting those found by "hand searching")

Vallejo-Torres et al. (2014)	2006-2010	England	Health concentration index.	Inequalities occur across the life- course but for some health issues there may be a period of equalisation in late adolescence and early adulthood.
Torre and Myrskylä (2014)	1975-2006	21 developed countries	Time series.	Income inequality is positively associated with mortality of males and females between the ages of 1 and 14 years and 15 and 49 years, and with mortality of females with ages of 65-89, albeit less strongly than for younger age groups.
Chauvel and Leist (2015)	2005, 2011	18 countries	Multilevel models.	Linear health gradients increase. Intergenerational transmission of status gains in importance in countries with higher income inequality.
Jutz (2015)	2008-2009	42 European countries	Two-step hierarchical estimation approach.	Income inequality has more impact on health inequalities than do social policies.
Lillard et al. (2015)	1913-2009, 1984- 2009	United States	Ordered probit models.	Exposure to income inequality in early life is related to worse health in later life.
Rambotti (2015)	1999	United States (plus international comparisons)	Bivariate and cross- sectional associations.	Poverty has a significant and adverse effect.

Among the most recent studies, there is also an interest in solving the apparent paradox that income appears to be related to health within countries but not between them. The explanation relies on the fact that in developed countries, which have already achieved a certain standard of living, increases in per capita GDP make little difference to the levels of health because of the *epidemiological transition*. (Mc Keown, 2009) that describes changing patterns of population age distributions, mortality, fertility, life expectancy, and causes of death. However, within countries, differences in living standards establish a social order in the population.

More recently, Pickett and Wilkinson (2015) have conducted a new review of the literature on the subject. Their work uses an epidemiological causal framework in order to infer the likelihood of a causal relationship between income inequality and health. They find a strong causal connection between income inequality and health, according to the exhaustive body of literature reviewed. IN the minority of studies that found no association, the following factors can be identified as problems: an inappropriate scale used to measure income inequality; the inclusion of mediating variables as controls; the use of subjective measures of health; and time periods that were too short. The authors also highlight that the effect of income inequality is to increase the gap between social classes or to widen differences in status.

Nevertheless, the real nature of the relationship between health and income is still not clearly defined, because of methodological issues. The literature that we have analysed raises a variety of questions about this relationship and shows the sensitivity in the different studies to the methodology used. The results depend to a great extent on the type of indicator used to measure health, the level of data aggregation and the causal effects among the variables.

Lastly, further research is necessary to investigate the role of income level, its composition and its distribution on health status and the labour market. To help with this, perhaps we can highlight the greater potential of individual studies, with the new databases available, for analysing hypotheses about a more detailed relationship between socioeconomic status, health and non-communicable diseases.

References

Chauvel, L., Leist, A.K. (2015) "Socioeconomic hierarchy and health gradient in Europe: the role of income inequality and of social origins", International Journal for Equity in Health, 14(1), p.132.

Eberstadt, N., Satel, S. (2004) Health and the income inequality hypothesis. A doctrine in search of data, The American Enterprise Institute Press, Washington, D.C.

Fuchs, V. (2004) "Reflections on the socio-economic correlates of health", Journal of Health Economics, 23, pp. 653-661.

Gravelle, H., Wildman, J., Sutton, M. (2002) "Income, income inequality and health: What can we learn from aggregate data?", Social Science and Medicine, 54, pp. 577-589.

Jutz, R. (2015) "The role of income inequality and social policies on income-related health inequalities in Europe", International Journal for Equity in Health, 14(1), p.117.

Lillard, D.R., Burkhauser, R.V., Hahn, M.H., Wilkins, R. (2015) "Does early-life income inequality predict self-reported health in later life? Evidence from the United States", Social Science & Medicine, 128, pp.347-355.

McKeown, R.E. (2009) "The epidemiologic transition: changing patterns of mortality and population dynamics", Americal Journal Lifestyle Med., 3(1 Suppl), pp. 19S-26S.

Pickett, K.E., Wilkinson, R.G. (2015) "Income inequality and health: A causal review", Social Science & Medicine, 128, pp. 316-326.

Rambotti, S. (2015) "Recalibrating the spirit level: An analysis of the interaction of income inequality and poverty and its effect on health", Social Science & Medicine, 139, pp. 123-131.

Torre, R., Myrskylä, M. (2014) "Income inequality and population health: An analysis of panel data for 21 developed countries, 1975-2006" Population Studies, 68, pp. 1-13.

Vallejo-Torres, L., Hale, D., Morris, S., Viner, R.M. (2014) "Income-related inequality in health and health-related behaviour: exploring the equalisation hypothesis", Journal of Epidemiology and Community Health, 68(7), pp.615-621.

Geospatial exposure analysis and citylevel modeling

The AIT geospatial model, developed under task WP5.2, aims at analyzing different aspects of health exposure at city-level, thus herein after referred to as *City-HeX*. Three case study areas were selected, representing one city each in three different European regions (Central-Eastern, Northern, Southern). Those cities are Lisbon (Portugal), Vienna (Austria), and Tallinn (Estonia).

Various environmental and socio-contextual risk factors are considered for health exposure analysis. These include directly measured parameters related to air pollution and temperature as well as approximated adverse influencing parameters like access to fast food and nightlife locations as well as green urban space. In a geospatial sense the aforementioned environmental variables refer to continuous fields (i.e. grids interpolated from point station measurement data) that are overlaid with disaggregated population distribution grids in order to directly estimate exposure patterns. The approximated variables refer to modeling exposure in a sense of people's accessibility to discrete spatial features such as fast food restaurants and urban parks.

Results of *City-HeX* serve as input for the micro-simulation model developed by OECD. While modeled in spatially-explicit manner (raster and vector), output is eventually aggregated and provided in tabular format (e.g. at district or municipality level) to OECD for integration in the micro-simulation modeling framework. For compliant interfacing of the two models, certain compromises need to be taken. Aside the tabular aggregation, the main issue to be addressed thereby is the handling of population and its characteristics. The OECD micro-simulation model handles population as individuals. *City-HeX*, in its initial setup, handles population in absolute terms, illustrating distribution patterns in space and time, without individual characteristics attached. That approach refers to the earlier-developed *DynaPop* model (Aubrecht et al. 2014) which represents dynamic spatio-temporal patterns of human activity (e.g. diurnal, weekly). To facilitate linking the two models, one option (keeping the dynamic setup) is to model population groups, e.g. certain age and activity groups (such as working population, students, retired people etc.). In order to be able to handle these population groups properly the temporal variation of population is limited to day-time and night-time representations.

In the first phase of WP 5.2 (July 2015 – Dec 2015) a first implementation and feasibility testing was performed for the Lisbon case study. This report covers the second phase (Jan 2016 – June 2016) where we concentrated on the modelling of population exposure for the Vienna case study.

Modelling of population exposure for Vienna case study:

The Vienna case study covers the administrative region of the city of Vienna: 23 districts covering 415 km2 with a population of 1.8 m people. The environmental parameters include PM_{10} and Ozone (O₃), the social parameters access to green urban areas, and to fast food locations. For temporal reference the years 2006 and 2013 were chosen due to data availability.

Modelling of night time population is based on census data, including five age classes (< 5, 5-15, 16-62, 65-84, >85), that were spatially disaggregated to a 100x100m raster representing residential areas and housing densities. For the day time population model commuting data were considered, as well as locations of workplaces, schools, kindergartens and retirements homes. An additional representation for summer (holiday) months was modelled taking into account less population (due to holiday absence) and outdoor locations (such as public outdoor pools).

Fig.1: Night-time (left) and daytime population (right) of the City of Vienna

For the modelling of PM_{10} and O_3 concentrations data from the air pollutant measurement network of the City of Vienna was used. For O_3 5 stations and for PM_{10} 13 stations are available that provide continuous measurement data over the last 10 years. Analysis of the data show the following trends: PM_{10} has higher concentrations in the winter half year, due to emissions of car traffic and heating, while O_3 has higher concentrations in summer due to stronger solar radiation (required for the formation of O_3).

Fig.2: Annual and seasonal means of PM₁₀ (left) and Ozone (right) for the City of Vienna

While annual/seasonal/monthly means might not be indicative for modelling local exposure, short term average values (1h/8h) were used according to WHO guidelines for Europe (WHO 2006). Days with exceeding loads (according to WHO recommendations) were selected and counted per measurements stations. These counts were then used for exposure mapping.

Fig.3: Days with exceeding loads for Ozone 2006 (left) and 2013 (right) for the City of Vienna

Modelling accessibility of population to certain locations is performed by adding up the population within a predefined distances to these locations. In order to do that exercise the relevant locations have to be represented in a spatial context. For the modelling of accessibility to green urban areas the urban atlas data of Vienna were used (<u>http://land.copernicus.eu/local/urban-atlas</u>). The distance to the class "green urban areas" were buffered and intersected with the population age groups resulting in number/percentage of people of a certain age class with access to parks within the given distance.

Fig. 4: Green urban areas (left) and access to green urban areas for population over 65 within 100m (right)

For the accessibility of fast food places data from Foursquare were used. Foursquare is a location based service for food venues and other places of interest (<u>https://foursquare.com/about</u>) providing not only location of venues but also a classification of the type of restaurant. The data set used included only locations that could be interpreted as fast food venues (such as Burger Joints, Hot Dog Joints or Pizza Places). Accessibility is calculated by defining a buffer around the locations and adding up the population per age group within the buffer area.

Fig. 5: Distribution of Fast Food venues (left) and access to Fast Food venues for pupils within 150m (right)

Next steps will concentrate on the analysis of population forecasts, expected immigration data and the future population pyramid for the city of Vienna as well as future trends in air pollution.

References

Aubrecht, C., K. Steinnocher, H. Huber (2014) DynaPop – Population distribution dynamics as basis for social impact evaluation in crisis management. ISCRAM 2014, 11th International Conference on Information Systems for Crisis Response and Management (pp. 319-323). University Park, PA, USA.

Aubrecht, C., K. Steinnocher (2016) Volunteered Geo-Dynamic Information for Health Exposure Assessment - a FRESHER Case Study. GI_Forum Journal for Geographic Information Science, accepted.

WHO (2006) WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. Global update 2005. Summary of risk assessment. World Health Organization, 2006.

