BD2Decide

Big Data and models for personalized Head and Neck Cancer Decision support

TITLE	Collected population data		
Deliverable No.	D7.2		
EDITOR	AOP - E. Martinelli		
Contributors	All Consortium partners		
WorkPackage No.	WP7	**WorkPackage Title**	Clinical validation studies
Status[1]	Final	**Version No.**	5
Delivery date	31/12/2016/2016	**Actual delivery**	17/01/2017
Dissemination level	PU		
DOCUMENT ID	D7.2 Collected population data		
FILE ID	BD2Decide D2.1		
Related documents	Technical Annex I DoA version 08/11/2016 - BD2Decide D2.1		

---

[1] Status values: TOC, DRAFT, FINAL

Distribution List

Organization	Name of recipients
AOP	T. Poli, E.M. Silini, E. Martinelli, G. Chiari, C. Caminiti, G. Maglietta, S. Rossi
VUMC	R. H. Brakenhoff, H. Berkhof, P. va de Ven, T. Klausch
UDUS	K. Scheckenbach, L. Colter
INT	F. Favales, L. Licitra, L. De Cecco, G. Calareso, G. Gatta, A. Trama, S. Canevari, G. Curzio
UPM	G. Fico, A. Fioravanti, L. Hernandez
Fraunhofer	F. Jung, S. Wesarg
ATC	V. Tountopoulos, E. Mantziou
AII	A. Algom, R. Shefi
POLIMI	L. Mainardi, V. Corino, E. Montin
MAASTRO	P. Lambin, F. Hoebers, A. Berlanga, R. Lejenaar
MME	S. Copelli, F. Mercalli
UNIPR	M. Vitale, G. Gobbi, M. Silva, E.M. Silini
European Commission	Project Officer: and all concerned E.C. appointed personnel and external experts

This document should be distributed as guidance to all the personnel of BD2Decide Consortium partners involved in the project execution.

Revision History

Revision no.	Date of Issue	Author(s)	Brief Description of Change
1	01.12.2016	E. Martinelli (AOP)	ToC and first contribution
2	19.12.2016	A. Fioravanti (UPM)	Added technical contribution
3	20.12.2016	A. Trama (INT)	Added contribution and revised sect. 1
4	30.12.2016	E. Martinelli (AOP)	Added contributions from INT and ISS
4.1	05.01.2017	E. Martinelli (AOP)	Added corrections by INT
4.2	12.01.2017	E. Martinelli (AOP)	Added corrections by ISS
4.3	16.01.2017	E. Martinelli (AOP)	Added final contribution by ISS
5	17.01.2017	E. Martinelli (AOP)	Final. Added contributions by UPM

Addressees of this document

This document is addressed to the BD2Decide Consortium. It describes the population data made available to the project for integration into models and big data analysis.

This document will be delivered to the European Commission.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Abbreviations and definitions

AII	All In Image
API	Application Programming Interfaces
ARPA	Regional Agency for Environmental Control (Italy)
ASCO	American Society of Clinical Oncology
BDA	Big Data Analytics
CDSS	Clinical Decision Support System
Csv	Comma separated value
DALY	Disability-Adjusted Life Year
DB	Database
EoD	End of Disease
HDC	Hospital Discharge Chart database: includes information on all discharges of Italian residents from hospitals of the Italian National Health System
EHF	Extremely High Frequency data (i.e. data collected by NHC on each patient's visit, treatment, pharma purchase etc.)
EHNS	The European Head and Neck Society
ESMO	The European Society for Medical Oncology
ESTRO	European SocieTy for Radiotherapy & Oncology
GHE	Global Health Estimates
H&NC	Head and Neck Cancer
IARC	International Agency for Research on Cancer
ISS	Istituto Superiore di Sanità (Italy Epidemiology National Institute)
INT	Istituto Nazionale dei Tumori
ISTAT	Italian Institute of Statistics
ISPRA	Istituto Superiore per la Protezione e la Ricerca Ambientale. It collects and publish data concerning environmental factors in Italy
NAS	Network-Attached Storage
NCCN	The National Comprehensive Cancer Network
NCR	National Cancer Registries. Collect individual data to define cancer incidence, mortality, prevalence and survival by type of cancers
RARECAREnet	Information network on rare cancers
RCE	Rare Cancers Europe (http://www.rarecancerseurope.org/)

TCIA	The Cancer Imaging Archive
VAT	Visual Analytics Tool
VUmc	VU Medical Center
WHO	World Health Organization
YLD	Years Lost due to Disability
YLL	Years of Life Lost

Executive summary

This document describes the population data that will be collected at National and EU level to complement patients' longitudinal data derived from data collection of the retrospective and prospective cohorts enrolled by BD2Decide clinical centres. These data are being collected and provided to technical partners - in particular statisticians and model builders at VU Medical Centre (VUmc) and Big Data analysis at All-In-Image (AII).

The above population data will be collected by IstitIstituto Superiore di Sanità (ISS, AOP subcontractor) at two levels:

- Aggregated population data i.e. data not directly linked to individuals but linked to clusters of populations (e.g. by territory, gender, localization of disease, census etc.): will be provided by end of December 2016, in csv form and in public databases directly accessible to BD2Decide Consortium. These data inform on the general health status, education, census, cancer incidence and prevalence, cancer death, treatment, presence of pollutants, risk factors (smoking, alcohol use/abuse) in sub-populations.

- High-resolution population data i.e. data linked to specific individuals (BD2Decide patients): a small cohort of Italian patients enrolled for BD2Decide who have provided consent will be selected to collect health related data from Italian Territory Health Agencies (e.g. medications, visits to specialized physicians, comorbidities, visits to GPs) when available. This high-resolution study will investigate how additional patient's data not collected during Head and Neck Cancer treatment may increase population profiling and patient's profiling and prognostic models accuracy.

A similar approach will be followed for data collected from cancer registries. Cancer Registries CRs) will be used grace to a collaboration with RARECAREnet EU project coordinated by the evaluative epidemiology unit of the Istituto Nazionale dei Tumori (INT). From CRs data INT will provide data at individual level with 2 different database:

- the database of a pilot study on centers of treatment for rare cancers. This pilot study was based on data already available at CR offices.

- the database of an high-resolution study on head and neck cancers. This high-resolution study was based on an ad hoc data collection by CRs.

The list of population data, already outlined in deliverable D2.1, is described in more details in this deliverable.

High-resolution data, also when linked to a specific patient, will be encoded using the BD2Decide patient ID (see deliverable D1.2 for a description of the coding procedure) prior to be stored into BD2Decide repositories.

About this document

This document is organized in three sections:

- Chapter 1 presents the population data provided to BD2Decide from cancer registries.

- Chapter 2 describes population data provided by AOP subcontractor ISS for sub-populations in Italy.

- Chapter 3 presents additional lifestyle, environmental, imaging and genomic data that will be made available to Big Data Analysis in aggregated form from public datasets.

- Chapter 4 describes how population data are managed.

- Chapter 5 summarizes the ethical and legal issues addressed for population data collection and usage in BD2Decide.

The annexes contain the structure of population data provided by INT, detaled description of population data provided by ISS along with the relevant metadata (structure).

1 BD2DECIDE CONSIDERED POPULATION DATA

1.1 CANCER REGISTRIES DATA

In order to complement the data collected from BD2Decide retrospective an prospective cohorts with a wider picture of Head and Neck Cancer epidemiology information in European populations, we will use Cancer Registries (CRs). CRs provide anonymized data at individual patient level and will be processed by INT to feed BD2Decide models recalibration and Big Data analytics tools. Two different inputs will be provided, constituting a Pilot Study (that will use only a subset of CR variables available in all considered CRs) and a High-resolution Study, that will consider a significant subset of more complete variables.

INT is in charge of the provision of these data.

1.1.1 Type of data

The population data provided for BD2Decide come from National Cancer Registries and regional cancer registries. These are organizations engaged in the gathering of information about all those sick with cancer **resident in a specified area (all new cases, any type of hospitals).**

Data sources are:

- Hospital discharge notes
- Anatomical and pathological cytology archives
- Clinical records
- Death certificates.

For BD2Decide INT will contribute with two different studies:

1. **Pilot study** (low resolution study). CRs used information already available in their database (possible only in CRs which routinely collect this information)

2. **High resolution study**. CRs went back to the different information sources to collect information defined in an ad hoc protocol.

1.1.2 Scope of data and added value for BD2Decide

The data collected by RARECAREnet contains clinical information including stage, age, site and sub-site of the tumors which are important prognostic variables. This will be useful to test the modeling and to integrated data coming from an unbiased population into the B2D data analytics.

1.1.3 Sources and selection criteria

Sources: population based cancer registries.

Criteria:

- High resolution study: CRs had to be national CRs or be regional but representative of geographical differences of head and neck incidence in the country.

- Pilot study: only national CRs.

1.1.4 Countries covered

- High resolution study: Ireland, Bulgaria, Netherland, Slovenia and Italy.

- Pilot study: Ireland, Bulgaria, Netherland, Belgium, and Slovenia.

1.1.5 High resolution study

RARECAREnet undertook an high resolution study in head and neck cancers. High resolution studies are observational study made in collaboration with population based CRs which go back to the clinical records and other information sources (such as discharge notes, pathological repots…) to collect details information not routinely available at CR office The head and neck cancer study was undertaken in 4 countries named: Bulgaria, Ireland, Italy, and Netherlands. We developed:

- common protocols,

- a MS ACCESS tool to collect the data and

- guidelines for the data collection.

A help desk, including clinical experts, was established to support cancer registry staff in the data collection.

The code book defined for the database of the high resolution study explains all the variables collected and the way they were analysed. The database has been created by INT and is available to BD2Decide.

Variable	Description		
country	country		
registry	registry		
id	Unique patient identification code		
gender	female/male		
dateofbirth	date of birth: dd/mm/yyyy (dd=15 if not allowed to reveal)		
dateofdiagnose	date of diagnose: dd/mm/yyyy (dd=15 if not allowed to reveal)		
primarytumoursitecode4digit	The numeric part of the ICD-O 3 topography code is reported (The "C" and "."is not included)		
morphologycode6digiticdo	ICD morphology code (6 digit)		
tnm_reference	edition of TNM manual used for codification:		
	U05	= UICC 5th edition	
	U06	= UICC 6th edition	
	U07	= UICC 7th edition	
stagect	Clinicical TNM, T primary site (for Bulgaria the cTNM was		

Variable	Description
	avialable only for patients not surgically treated)
stagecn	Clinicical TNM, N lymph nodes
stagecm	Clinicical TNM, M metastases
stagept	Pathological (after surgery) TNM, T primary site
stagepn	Pathological TNM, N lymph nodes
stagepm	Pathological TNM, M metastases
ctscandone	was the CT scan done?
	• 0 = no
	• 1 = yes
	• 2 = no information available (In Bulgaria CT scan has to be paid and is expensive thus, it is very likely that no information is due to the fact that other diagnostic imaging such as echography were performed. Netherlands collected this information only for a sample. The variable selim identifies cases included in this sample)
mriscandone	was the MRI scan done?
	• 0 = no
	• 1 = yes
	• 2 = no information available (In Bulgaria CT scan has to be paid and is expensive thus, it is very likely that no information is due to the fact that other diagnostic imaging such as echography were performed. Netherlands collected this information only for a sample. The variable selim identifies cases included in this sample)
hospitalofdiagnose	Unique hospital identification code or name of the hospital. The hospital of diagnose was the hospital where the pathological examination was performed or requested
surgeryofprimarytumourdone	was the surgery done?
	• 0 = no
	• 1 = yes
	• 2 = no information available
dateofsurgery	Date of surgery (if done):
	dd/mm/yyyy (dd=15 if not allowed to reveal)
dateofsurgmiss	date of surgery missing: 1=yes
hospitalofsurgery	Hospital where surgery was done. Unique hospital identification code or name of the hospital.
resectionstatusr0r1r2	Margins after surgery:
	• R0 (tumour free margin)
	• R close (inferior 5mm)
	• R1
	• R2
	• R+
	• not included in the pathological report (this detail is

Variable	Description
	available only for Italian cancer registries)
	• no information available
informationontechniqueusedforsurgery	information on the technique used for surgery:
	surgical resection. Only for Italian cancer registries we have information on whether radical or local defined as follows: radical resection: if T1, the surgery is radical if focused on T; if stage 3 e 4, the surgery is radical if it is focused on T and N; It is always radical if it includes the N regardless of T (T1,T2,T3)
	local excision (any T without an excision of N)
	other (cryotherapy photodynamic therapy, electrocautery procedure, cryosurgery, lasertherapy, thermo-ablation)
	palliative (this was specified by Bulgarian cancer registry only)
	no information available
information available in the post surgery pathologial report (Netherlands and Ireland collected this information for a sample of cases identified by the variable selpat)	
noinformationavailable	the information is not available? 0= the information is available / 1= the information is not available
siteandlateralityofthecarcinoma	Availability of site and laterality of the carcinoma in the post surgery pathological report (true=1 /false=0)
maximumdiameteroftumour	Availability of maximum diameter of tumour in the post surgery pathological report (true=1 /false=0)
maximumdepthofinvasion	Availability of maximum depth of invasion in the post surgery pathological report(true=1 /false=0)
histologicaltypeofcarcinoma	Availability of histological type of carcinoma in the post surgery pathological report (true=1 /false=0)
degreeofdifferentiationgrade	Availability of degree of differentiation in the post surgery pathological report (grade) (true=1 /false=0)
patternofinvasion	Availability of pattern of invasion in the post surgery pathologial report (true=1 /false=0)
evidenceofhumanpapillomavirusinf	Availability of information about human papilloma virus infection in the post surgery pathologial report (true=1 /false=0)
marginstatusr0rcloseinferior5mmr	Availability of margin status in the post surgery pathologial report (true=1 /false=0)
lymphnodeinvolvement	Availability of lymphnode involvement in the post surgery pathologial report (true=1 /false=0)
chemotherapyisused	was chemotherapy done?
	• 0 = no
	• 1 = yes
	• 2 = no information available
dateofstartchemotherapy	Date of start of chemotherapy (if done):
	dd/mm/yyyy (dd=15 if not allowed to reveal)
dateofchemomiss	was the date of start of chemotherapy missing? (1=yes)
hospitalofchemotherapy	Unique hospital identification code or name of the hospital. Hospital where the chemotherapy was administered
chemiotherapy_2isused	was a second chemotherapy done?
	• 0 = no
	• 1 = yes

Variable	Description
	• 2 = no information available
dateofchemotherapy_2	Date of start of second chemotherapy (if done):
	dd/mm/yyyy (dd=15 if not allowed to reveal)
dateofchemo2miss	was the date of start of second chemotherapy missing? (1=yes)
hospitalofchemotherapy_2	Hospital where the second chemotherapy was administered. Unique hospital identification code or name of the hospital.
radiotherapyisused	Was radiotherapy done?
	• 0 = no
	• 1 = yes
	• 2 = no information available
dateofstartradiotherapy	Date of start of radiotherapy (if done):
	dd/mm/yyyy (dd=15 if not allowed to reveal)
dateofradiomiss	was the date of start of radiotherapy missing? (1 =yes)
hospitalofradiotherapy	Hospital where the radiotherapy was administered. Unique hospital identification code or name of the hospital.
informationontypeofradiotherapy	Information on type of radiotherapy:
	no information available
	brachytherapy
	external radiotherapy NOS
	IMRT (intensity modulated radiotherapy)
	3DCRT (3D conformal radiotherapy)
	proton therapy
	other radiotherapy (please specify)
ifotherradiotherapypleasespecify	Other kind of radiotherapy
lastdateoffollowup	Date of death or last follow-up dd/mm/yyyy (dd=00 if not allowed to reveal)
statuspatient	Status of patient:
	• 0 = Alive
	• 1 = Dead
	• 2 = Lost to follow-up
site	tumour site:
	hypopharynx
	larynx
	oral_cavity
	oropharynx
selpat	selection of patients in Netherland and Ireland for which information available in the pathological report were collected. The data collection was intentionally performed on a sample of cases
selim	selection of patients in Netherland for which we have information about the imaging. The data collection was intentionally performed on a sample of cases
stcond4	stage: early/advanced/M (metastatic)/missing (assuming MX=M0 and NX or N missing=N0 only for larynx tumors with pT<2 and

Variable	Description
	pN0)
treatment5	we have generated a treatment variable defining adjuvant therapy on the basis of strictly time discussed with our head and neck experts. • no treatment • only chemotherapy (if started within max 3 months from diagnosis) • only radiotherapy (if started within max 3 months from diagnosis) • only surgery (if started within max 3 months from diagnosis) • concomitant chemo-radio (concomitant if chemotherapy starts from 1 day before the radiotherapy or max 49 days after the radiotherapy) • surgery + chemotherapy (if chemotherapy started max within 2 months from surgery date) • chemotherapy + surgery (if surgery started max within 3 months from chemotherapy) • surgery + radio or + concomitant radio-chemotherapy (concomitant if chemotherapy and radiotherapy start the same day or chemotherapy starts from 1 day before the radiotherapy and 35 days after the start of radiotherapy) • chemotherapy + radiotherapy (if radiotherapy started max within 3 months from chemotherapy) • radiotherapy+ chemotherapy (if chemotherapy started max within 3 months from radiotherapy) • combined treatments beyond 3 months (surgery, radio, chemo administered beyond 3 months from each other thus without adjuvant intent) • missing treatment (there could be some information but it is not enough to define the treatment) • information available on surgery and radiotherapy (only for Bulgaria to reduce the missing due to lack of date of start treatment) • information available on surgery and chemotherapy (only for Bulgaria to reduce the missing due to lack of date of start treatment)
surgerydone3	This variable is different from the "surgeryofprimarytumourdone" because it is derived from the treatment5 • 0 = no • 1 = yes • 2 = the value of variable treatment5 is "missingtreatment"
radiodone3	This variable is different from the "radiotherapyisused" because it is derived from the treatment5 • 0 = no • 1 = yes • 2 = the value of variable treatment5 is

Variable	Description
	"missingtreatment"
chemodone3	This variable is different from the "chemotherapyisused" because it is derived from the treatment5 • 0 = no • 1 = yes • 2 = the value of variable treatment5 is "missingtreatment"
prok	number of information available in the pathological report (99 missing)
prokninf	number of information available in the pathological report for patients without lymph node involvement (99 missing)

<div align="center">

Table 1-1. Data currently available in the high resolution DB

</div>

The High resolution Study can provide 9.436 patients' records, of which 4.111 (43,57 %) are of interest for BD2Decide (Stage III and IV Head and Neck cancer cases).

Notably also the data of earlier stage cases are of interest to us as they may describe/highlight different distribution by stage incidence/prevalence/death in different regions.

Countries	Head and Neck cancers	
	All stage	**Advanced stage**
Bulgaria	972	89**
Ireland	1.323	640
Netherlands	6.185	3.000
Slovenia (only larynx)	219	90
Italy *	737	292
Total	**9.436**	**4.111**

<div align="center">

Table 1-2. Number of patients records availably by Country for the High resolution Study

* 9 cancer registries; ** % missing very high

</div>

1.1.6 Pilot study

RARECAREnet undertook also a study to understand whether rare cancers treatment is centralised or widespread across several hospitals. Five countries contributed to this pilot study: Belgium, Bulgaria, Finland, Ireland, Netherlands. RARECAREnet conducted 2 meetings with CRs to discuss the objectives of the pilot study and to define a common protocol for the data collection. Information were collected for all 198 rare cancers (including head and neck cancers) on date and place of diagnosis, data and place of treatments (surgery, radiotherapy, chemotherapy, others), stage at diagnosis, date and status of patients (alive or dead) at follow-up. Data were analysed in order to develop indicators of centralisation and of dispersion of treatment. In addition, we defined the hospital volume of hospital of treatment and assessed for some rare cancers the relationship between hospital volume and outcome.

The following data are collected and will be transferred to BD2Decide:

Patient data	Diagnosis data	Treatment data (5 treatments collected)	Follow-up data status
Birth date	Date	1. date, type, hospital, hospital type	vital status
Sex	Hospital	2. date, type, hospital, hospital type	closure date (date of last visit)
	Topography	3. date, type, hospital, hospital type	
	Morphology	4. date, type, hospital, hospital type	
	Stage TNM	5. date, type, hospital, hospital type	
	Stage EoD	6. date, type, hospital, hospital type	
		7. date, type, hospital, hospital type	

Table 1-3 Pilot Volume Study: data collected

We have selected the cases fulfilling BD2Decide inclusion/exclusion criteria, namely cases with Head and Neck cancer localized in the following regions: larynx, oral cavity, oropharynx, hypopharynx, as shown in the following table. We note that Advanced stage patients (Stage III and IV) constitute 40,9% of total cases registered in CRs.

Country	Year of diagnosis	No cases (all stages)	No cases advanced
Belgium	2004-2007	7918	3013
Bulgaria	2005-2007	2938	1161
Ireland	2000-2007	2498	943
Netherlands	2005-2007	6075	2297
Slovenia	2000-2007	2874	1505
Total		**22.303**	**9.120**

1.1.7 Current status

The database of the high resolution study and of the pilot study are ready. RARECAREnet is seeking the permission to use the data for B2D decide. We requested to all CRs to update the follow-up to estimate 5 year RS (Slovenia Sept 2016; Netherlands Feb 2016; Italian CRs most 2015 some 2014). For the high resolution study, nine out of 10 Italian CRs (discussion on going with Friuli CR) and Bulgaria CR agreed to collaborate with B2Decide. Waiting for positive response from Slovenia, The Netherlands, Ireland. For the Pilot Study we already received permission to use the data from Belgium, and Bulgaria. Ireland, Slovenia, and Netherlands are still pending.

2   POPULATION DATA FOR ITALIAN PATIENTS

Population demographic, lifestyle and socioeconomic status data are collected from Italian Populations. This will allow to identify (as compared to Germany and the Netherlands, where such wealth of population data is not collected in the frame of this project) how much this population-related data can contribute to refine prognostic models and orient treatment.

2.1   BACKGROUND

Population data are provided for Italy in order to describe different aspects of the Italian context (background), in particular there are seven areas: 1) demographic and socio-economics; 2) mortality (overall and cause-specific); 3) morbidity, 4) disability and hospital discharges; 5) health care resources; 6) health care utilization and costs; 7) lifestyles.

Population data (i.e. mortality) and indicators derived from data (i.e. life expectancy) are available and publicly accessible in grouped format and do not need any informed consent from citizens.

In BD2Decide, population data will contribute to refine prognostic models and orient treatment.

Population data are collected by different institutions in Italy: the Italian National Institute of Statistics (ISTAT), the Italian Ministry of Health, the Italian Medicines Agency (AIFA), the Italian National Institute for Environmental Protection and Research (ISPRA), and the Italian National Health Institute (ISS). The data are collected for administrative purposes (as is the case for the hospital discharge database) or for surveillance purposes (as is the case for the PASSI study from ISS).

Data collection may be routinely (i.e. mortality statistics collected every year) or sporadic (i.e. surveys as the Italian Multipurpose survey from ISTAT), data may refer to the whole Italian population (i.e. Hospital Discharge Cards database-HDC, Outpatients files, Pharmaceutical records) or to a sample representative of the Italian population (the Italian Multipurpose survey from ISTAT, PASSI study from ISS, …).

Population data collection in Italy is largely geographically heterogeneous in terms of population coverage, time series availability and area of reference of the data/indicators (i.e. census area, province, Region).

2.2   SELECTION CRITERIA

Selection criteria of the population data and indicators are:

- The geographical coverage: the whole country or samples representative of the whole country

- The time series availability: at least two years of data available within the period of interest of the clinical cohorts (2008-2014)

- The level of information: regional or sub-regional areas (i.e. province, census area)

- The accessibility (necessary informed consent when applicable): publicly available data

- The quality of the information: structured data from administrative data sources or surveillance data sources

- The added value of the information: no a priori hypotheses, all information useful to describe the life context of the patients cohorts are collected.

2.3 LEVEL OF DATA AND STUDIES EXECUTED

The data collected by ISS on Italian population cover demographic, socio-economic and health-related status of the population. They are of two types:

1. **Aggregated data**: they describe the average population by region (geographic area) age, sex and - in some cases - education and census. These data are anonymous and grouped, they are publicly available and do not need any informed consent from citizens, being part of Institutional National demographic researches.

 Inclusion criteria for these data are:

- available at least at regional level (data available at national level only have been excluded)

- available at least for 2 consecutive years in the period 2008-2014 (data/indicators available for one year only or for a period outside the reference timeframe 2008-2014 have been excluded).

- The access to these data requires the knowledge of the region of residence of the involved patients, in order to require the data to the relevant administration.

2. **High-resolution data**: a sub-study to be investigated and performed if possible, with the aim to collect personalized socioeconomic and health-related information for a patient-specific characterization. During a preliminary meeting held in Parma in March 2016 ISS and AOP agreed to identify a sub-population of enrolled cases in Italy, from those Italian Regions where electronic health information could be available upon request. For these data ISS will perform a *feasibility study* in order to determine the possibility to access these data and the availability of such data on the territory and for the time periods covered by BD2Decide patients' enrolment.

If possible and ethically and legally feasible, a high-resolution study will be executed on the selected subgroup of Italian patients, linking their personal socioeconomic and health-related data from administrative DB to BD2Decide cancer-related information. Data will be collected (under informed consent by both the patient and the relevant health care institution in charge of patient's data custody), then anonymized and linked through the BD2Decide ID to the encoded patient record in BD2Decide. If available a sub-study might be performed (although not indicated in the DoAThese data are used to complement data collected for the patient by the cancer treatment centre.

Individual patient's data are linkable (at least at theoretical level) to the patients' from Italian retrospective and prospective cohorts.

Source of data

The high-resolution data come from Italian administrative DB (Mostly from the Ministry of Health and Regional Information Systems).

Inclusion criteria:

- available for the individual patient

- available at least for 2 consecutive years in the period 2008-2014 (data/indicators available for one year only or for a period outside the reference timeframe 2008-2014 have been excluded).

. These data would need patients' consent to access, either already given to the relevant administrations or specifically requested to patients by the Italian hospitals (AOP, INT) participating to BD2Decide.

Characteristics of aggregated data

Aggregated data have the advantage of being publicly available and as such they do not need patients' informed consent for access and use. They can be used for tuning the prognostic models and for Big Data Analysis to identify population-related indicators, i.e. indicators derived from the average characteristics of the populations based on Country/Region of residence. Data are provided in tabular format as described in annex. ISS is in charge of the identification of available data of interest for BD2Decide (which data, aggregation level, time periods, covered areas, data sources), to check data quality and to provide the available data along with a metadata description and data structure.

The data provided have the following characteristics:

- Largely regionally heterogeneous

- Different ways to collect the data depending on the source (surveillance, administrative, …)

- Population based (CRs, HDC, Outpatients files …) vs samples (Multipurpose survey from ISTAT, Passi Study …)

- Individual level (CRs, HDC, …) vs aggregated level (census data, some survey data …)

- Routinely collection versus sporadic.

Time series availability

- EHF started in 2015,

- CRs are active since 1976,

- HDC are available in electronic form since 2001

The majority of data come from Italian Institute of Statistics (ISTAT) and can be queried on-

line and downloaded from the web site: www.dati.istat.it (Italian) and http://dati.istat.it/?lang=en&SubSessionId=1cc72e44-e359-495d-8bd9-bb5db11836b9&themetreeid=-200 (English).

Characteristics of high-resolution data

The data provided have the following characteristics:

- Available and specific for the individual patient
- Different ways to collect the data depending on the source and the patient's residence
- Linked to each specific patient or to patients' homogeneous subgroups
- Routinely collection for patients' related events.

2.3.1   Scope of data, sources and added value for BD2Decide

As already indicated in deliverable D2.1, these data are aggregated (no individual citizens' records available) and cover the following domains:

1. Demographic and Lifestyle behaviors data (available for years sequences starting from year 2006):

   - Demographic indicators such as: mortality and fertility rate, life expectancy, dependency rate (population 0-14 and 65+ divided by population 15-64), elderly rate by (census area or city of residence) at population level from the Italian National Institute of Statistics (http://demo.istat.it/).

   - Lifestyle risk factors and preventive measures such as smoking habits, alcohol consumption, dietary habits, overweight, physical activity, vaccinations, preventive health screenings aggregated by health local unit from surveys administered to representative samples of the Italian population (for example, the PASSI study: http://www.epicentro.iss.it/passi/) and from ISTAT and ISS records.

2. Health data of the population and access to medication: these data will be used in aggregated format and - upon consent and request to the relevant local Health Units) also for a sub-study on a selected cohort of Italian patients from Parma and INT. These cases will be separately studied to identify how health data not available at the hospital treating the patients for Head and Neck cancer (for privacy reasons), may influence the prognosis.

   - Hospital Discharge Cards (HDC) database (from 2001 to 2014): clinical variables (main and secondary diagnoses, main and secondary treatments); administrative variables (type of admission: ordinary or day hospital, date of admission and discharge, diagnostic related group, length of stay, cost of the treatment …). These data are available at aggregated and disaggregated level from Italian Ministry of Health (http://www.salute.gov.it/portale/temi/p2_4.jsp?area=ricoveriOspedalieri).

   - Drug prescriptions: data at aggregated levels will be collected from Regional Health Systems that manage such data in electronic format.

- Outpatient records: these data will be used only for the sub-study. They include information on all treatments and procedures administered from local ambulatories of the Italian Public Health System. They are being requested to the relevant territorial/local Health Agencies (AUSL).

REGIONS GEOGRAPHIC AREAS TYPES OF MUNICIPALITY	Drinks non-alcoholic aperitifs	Drinks non alcoholic aperitifs exceptionally	Drinks alcoholic aperitifs	Drinks alcoholic aperitifs exceptionally	Consuma amari Drinks bitter liquors	Drinks bitter liquors exceptionally	Drinks liquors	Drinks liquors exceptionally	Drinks alcohols outside meals	Drinks alcohols outside meals less than once a week	Does not drink wine, beer or alcohols outside meals	Population
Piemonte	2.033	1.211	1.392	826	1.069	652	990	709	1.273	989	1.254	3.972
Valle d'Aosta/Vallée d'Aoste	63	35	48	24	29	17	33	21	47	32	36	114
Liguria	684	417	506	297	348	196	381	235	407	326	502	1.443
Lombardia	4.300	2.508	3.222	1.688	2.331	1.379	2.364	1.504	2.948	2.203	2.918	8.908
Trentino-Alto Adige	437	264	364	188	258	158	278	196	416	317	303	923
Bolzano/Bozen	*210*	*120*	*191*	*96*	*154*	*83*	*160*	*110*	*218*	*173*	*132*	*451*
Trento	*227*	*144*	*173*	*92*	*104*	*75*	*118*	*85*	*197*	*144*	*171*	*472*
Veneto	2.196	1.239	1.698	820	1.207	792	1.245	859	1.639	1.158	1.403	4.403
Friuli-Venezia Giulia	526	296	404	199	288	189	308	220	442	311	371	1.103
Emilia-Romagna	1.687	1.004	1.486	782	1.068	634	1.200	760	1.327	994	1.201	3.981
Toscana	1.441	923	1.145	687	826	548	942	625	960	778	1.015	3.375
Umbria	352	207	263	158	202	129	199	131	199	147	271	804
Marche	588	359	481	250	336	213	371	251	386	294	437	1.396
Lazio	2.397	1.425	1.682	1.049	1.447	996	1.295	888	1.153	876	1.972	5.242
Abruzzo	608	313	412	218	348	204	256	166	353	268	385	1.204
Molise	152	67	96	47	85	45	63	36	83	62	112	286
Campania	2.570	1.498	1.471	902	1.555	990	1.067	720	946	735	2.172	5.202
Puglia	1.703	1.070	956	605	1.188	718	749	515	714	564	1.375	3.680
Basilicata	247	126	140	84	133	87	91	58	114	89	203	523
Calabria	906	532	556	361	594	366	388	270	406	320	664	1.771
Sicilia	2.127	1.292	1.314	806	1.297	777	866	547	875	698	1.974	4.545
Sardegna	746	402	458	283	391	237	391	272	508	375	572	1.512
North west Italy	7.080	4.170	5.169	2.835	3.776	2.245	3.768	2.470	4.675	3.550	4.709	14.438
North-east Italy	4.847	2.802	3.951	1.989	2.821	1.773	3.031	2.036	3.824	2.780	3.278	10.411
Central Italy	4.777	2.914	3.570	2.145	2.812	1.886	2.808	1.895	2.697	2.095	3.694	10.817
South Italy	6.185	3.606	3.630	2.217	3.903	2.411	2.615	1.765	2.616	2.038	4.911	12.667
Italian Islands	2.873	1.693	1.772	1.089	1.687	1.014	1.257	818	1.383	1.073	2.546	6.057
Metropolitan municipality	4.007	2.443	2.916	1.747	2.485	1.563	2.250	1.505	2.275	1.769	2.928	8.280
Suburbs of metropolitan area	3.172	1.841	2.310	1.280	1.954	1.174	1.750	1.101	1.862	1.403	2.225	6.665
Municipality with up to 2.000 inhabitants	1.535	807	1.027	537	813	489	727	481	1.036	740	1.071	2.990
Municipality with 2001-10000 inhabitants	5.804	3.374	4.127	2.291	3.276	2.013	3.051	1.987	3.754	2.779	4.459	12.626
Between 10.001-50.000 inhabitants	6.760	4.018	4.590	2.546	3.811	2.374	3.360	2.270	3.778	2.870	5.322	14.508
More than 50.000 inhabitants	4.484	2.702	3.123	1.872	2.660	1.714	2.340	1.639	2.491	1.974	3.132	9.320
Italy	**25.763**	**15.185**	**18.092**	**10.275**	**14.999**	**9.328**	**13.479**	**8.984**	**15.196**	**11.535**	**19.138**	**54.389**

Table 2-1. Example population data: lifestyle habits and risk factors by Region (Italy, 2015)

2.3.2 Population health-related data

We will consider the following data:

- Lifestyle behaviors: smoking habits, alcohol consumption (beer, wine, alcoholic drinks).

- Health status: healthy, suffering from chronic diseases.

- Use of health services: hospitalization, hospitalization days, access to GPs

These data can be detailed by region, age range, education level, type of employment, sex and combination of such indicators

2.3.3   Selection criteria

To integrate the models and big data analysis with external data have evaluated:

- Geographical coverage.

- Time series availability.

- Level of information.

- Accessibility (necessary informed consent when applicable).

- Quality of the information.

- Added value of the information.

2.3.4   Current status

Aggregated data have been collected and are being provided to technical partners for use in models recalibration. Data will be fully available by 31/12/2016.

Permissions for the high resolution study are being requested for subsets of patients from Parma hospital. In annex to this document we present a list of provided variables and datasets.

2.3.5   Levels of data

Population data/indicators are described according to the following three levels:

1. First level: the areas of the Italian background.
2. Second level: the data/indicators included in each area.
3. Third level: the data/indicator features, including type of data source (institution collecting the data; sample vs population; administrative vs surveillance data source) and data availability (geographical coverage, level of aggregation, time series, other stratification variables).

First level: background areas

There are seven areas here considered to describe the Italian background (see Figure 1):

- Demographic and socio-economic area;
- Mortality-based
- Morbidity, disability and hospital discharges/activity
- Health care resources
- Health care utilization and costs
- Lifestyles
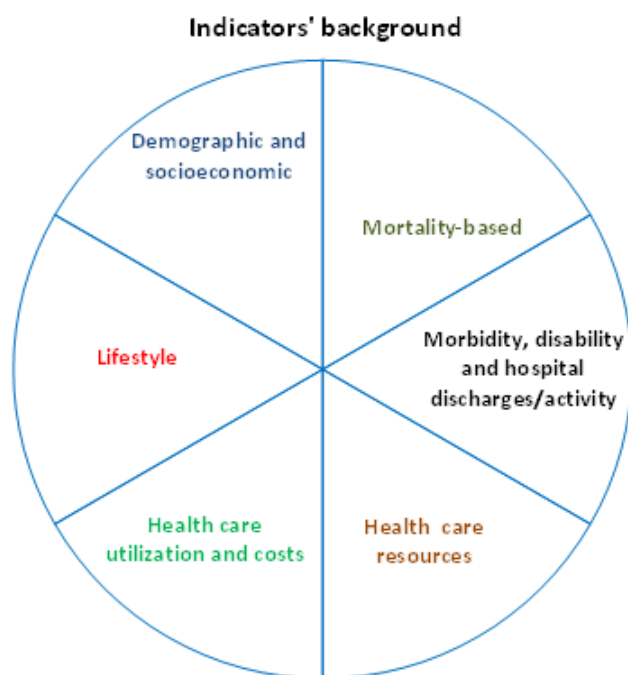
Figure 2-1. Italian background areas

Second level:

At this level are listed all data/indicators included in each area of the Italian background.

- Demographic and socio-economic area: includes all aspects related to the dimension (resident population) and the age structure of the population (dependency ratio, ageing rate, population aged 65+, 85+), to the socioeconomic condition (foreign resident population, education, occupation, poverty) and to the demographic dynamic of the population (fertility rate, birth rate, mother's mean age, abortion);
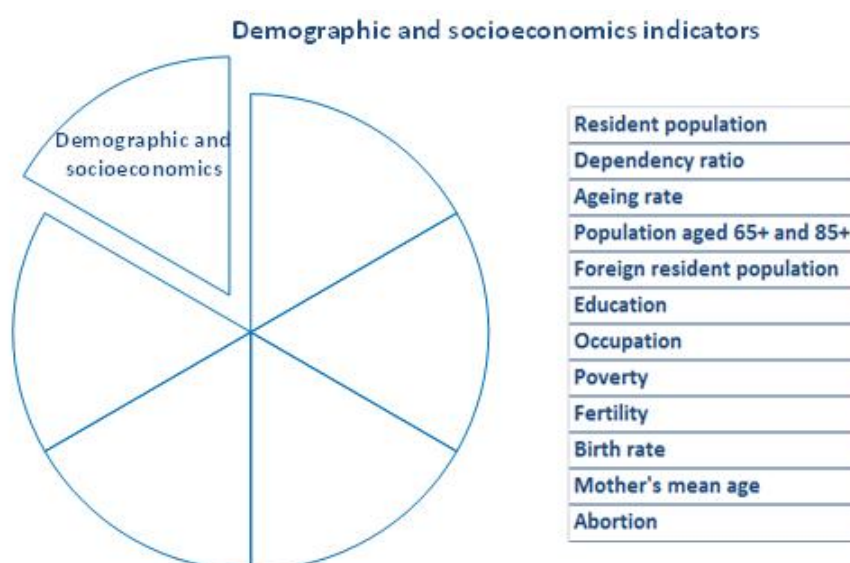
Figure 2-2. Demogpaphyc and socio-economic background

- Mortality-based: includes all aspects related to ageing and future population projections (life expectancy at birth, healthy life expectancy, Infant deaths, stillbirths, enanatal deaths, general mortality, cause-specific mortality)
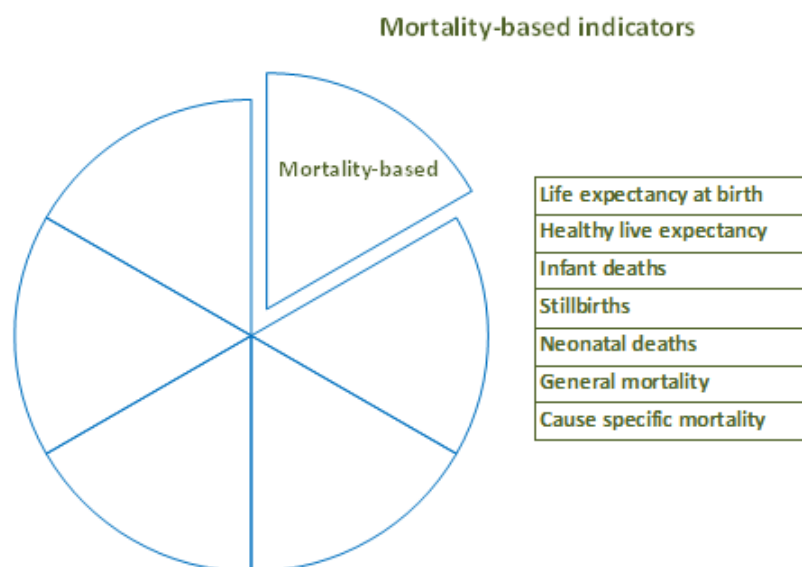
Figure 2-3. Mortality indicators

- Morbidity, disability and hospital discharges/activity: includes incidence and prevention related to infectious diseases (vaccination, infectious and chronic diseases), burden of non infectious diseases (incidence and prevalence of malignant tumours and chronic diseases).
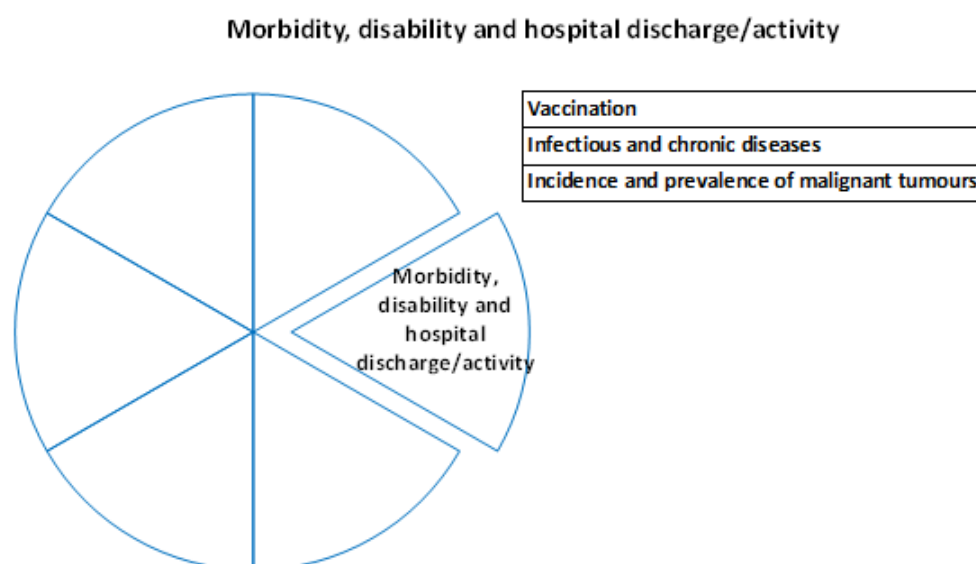
Figure 2-4. Health-related population data

- Health care utilization and costs: includes aspects related to hospital admissions and discharges, general practitioners, public spending and drugs consumption.

Health care utilization and costs indicators

Hospitalizations and discharges
GNP
Public spending
Drugs consumption

Healtc care utilization and costs
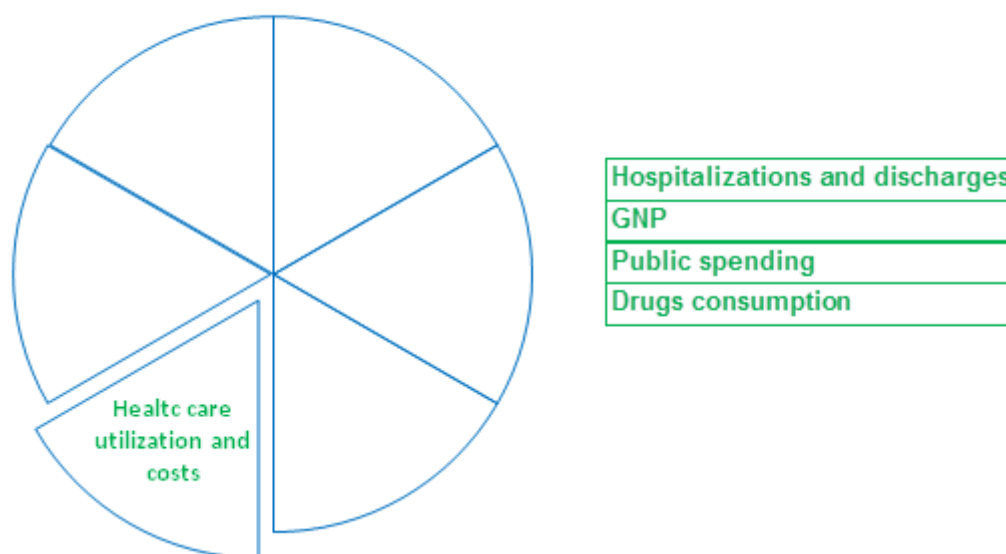
Figure 2-5. Use of healthcare resources by citizens

- Health care resources: includes the availability in the territory of general and health care specific infrastructures (i.e. oncological health care), number of available beds (in-hospital beds, day hospital beds, physical therapy beds), hospital characteristics

Health care resources indicators

General and oncological health care
Hospital, day-hospital and physical therapy beds
Hospital characteristics

Healtc care resources
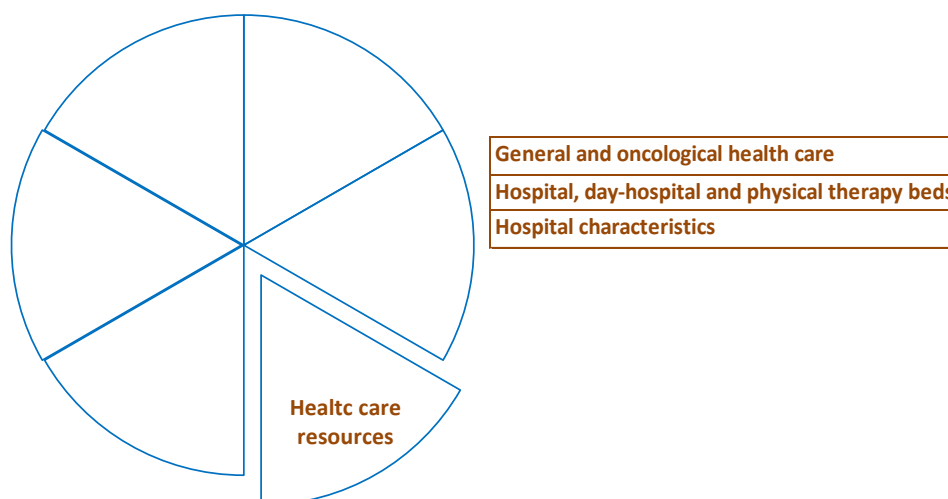
Figure 2-6. Indicators of distribution of healthcare facilities in Italian terriories

- Lifestyles: includes risk factors exposure and benefits related to lifestyles (anthropometric data, smoking habits, physical activity and sports, eating habits)

Lifestyle indicators
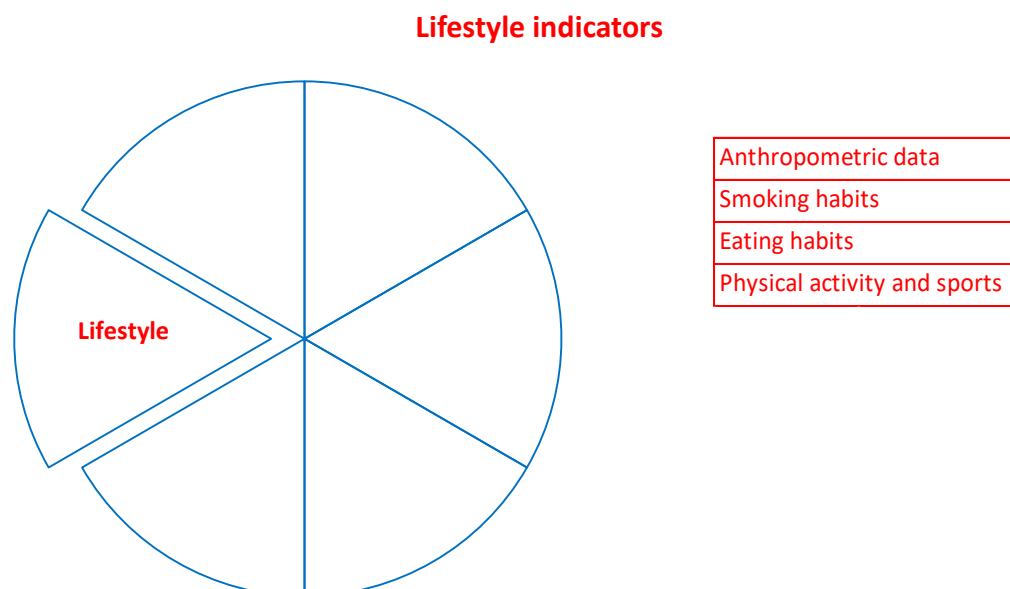
Figure 2-7. Lifestyle indicators at populationn level

Third level:

At this level for each data/indicator included in each area of the background there is a description of specific features concerning: the formula used to calculate the indicator, the stratification variables (age, gender, …), the data availability (geographical coverage, level of aggregation, time series), the data source (institution collecting the data; sample vs population; administrative vs surveillance data source).

1. **Demographic and socioeconomic indicators**

- Resident population

- Dependency ratio

- Ageing rate (AGrate)

- Population aged 65+; 85+

- Foreign residents at January 1st

- Education level

- Occupation: activity rates, employment, unemployment,

- Poverty: Incidence of poor persons

- Health demographics indicators:

    o Fertility rate

    o Birth rate

    o Mother's mean age at birth

    o   Abortion, Mother's mean age at sp.abortion

2. <u>Mortality based indicators</u>

- Life expectancy at age x

- Infant deaths and mortality rate

- Total mortality, cause-specific mortality, mortality for drugs1, alcohol

3. <u>Morbidity, disability and hospital discharges/activity</u>

- Flu vaccination

- Infectious and Chronic diseases

- Self-reported health status

- Incidence and prevalence of malignant tumours

4. <u>Health care resources</u>

- General health care practitioners (GHP)

- Medical oncology doctors

- Outpatients and residential facilities

- Hospital, day-hospital and physical therapy beds

5. <u>Health care utilization and costs</u>

- Hospitalizations and discharges

- Length of stay of Acute disease patients outside the region of residence

- Drugs consumption (covered by NHS)

- Financial resources
  - o GDP
  - o Public health spending

6. <u>Lifestyle</u>

- Obesity, overweight

- Smoking

- Eating habits

- Alcohol consumption

- Physical activity and sports

Detailed description is provided in Annex.

3 EXTERNAL DATA FROM OTHER SOURCES RELEVANT FOR BD2DECIDE

3.1.1 Epidemiology data

Main sources for epidemiology data are International Agency for Research on Cancer (IARC) databases and World Health Organization (WHO) databases.

For model recalibration VUmc statisticians may use additional epidemiology information provided by WHO that can be extracted from WHO Databases as required. Data are aggregated by Country. An example of mortality data series by Country for Head and Neck cancer is shown in the table below.

Years	Belgium	Germany	Italy	Netherlands	Poland	Portugal	Romania	Slovenia
2015				654			2174	146
2014		5448		615	2675	694	2158	
2013	595	5473		646	2465	696	2144	152
2012	581	5393	2843	629	2376	762	2151	152
2011	520	5386	2859	628	2239	764	2055	163
2010	521	5020	2745	609	2227	690	1976	169
2009	521	4982	2687	630	2186	702	1978	158
2008	537	4946	2755	557	2168	695	1953	164
2007	487	4777	2801	535	2110	640	1885	151
2006	515	4734	2745	554	1979		1834	148
2005	527	4819		538	1886		1745	131
2004	501	4444		569	1872		1743	110
2003	499	4786	2746	566	1857	609	1634	131
2002	478	4699	2860	547	1762	607	1735	147
2001	513	4853	2886	508	1767	585	1635	179
2000	487	4645	2866	533	1690	559	1488	127
1999	519	4941	2943	519	1685	541	1367	134
1998	566	4965	2943	477		535	1335	169
1997	575	4778	2871	459		572	1285	158
1996	532	4766	2939	461	1673	506	1284	158
1995	552	4684	2949	452	1760	483	1144	143

Table 3-1. Example: No of deaths. Cancer of lip, oral cavity and pharynx, both sexes

(source WHO DB http://apps.who.int/healthinfo/statistics/mortality/whodpms/)

The IARC is the most relevant institution for epidemiology study in EuropeData can be downloaded here: http://globocan.iarc.fr/Pages/online.aspx. In the web site @Cancer Incidence in 5 Continents" you can find the following incidence time trend graphs only for The Netherlands and Slovenia, unfortunately the site selection is fixed and some countries are missing. Here below the relevant links:

- Netherlands

    o Oral cavity and pharynx:
      http://ci5.iarc.fr/CI5plus/old/Graph4p.asp?cancer%5B%5D=20&male=1&female=2&country%5B%5D=5280099&sYear=1950&eYear=2010&stat=3&age_from=4&age_to=18&orientation=1&window=1&grid=1&line=2&moving=1&scale=0&submit=%C2%A0%C2%A0%C2%A0Execute%C2%A0%C2%A0%C2%A0

    o Larynx: data can be extracted by EU Country and region and exported in tables or visualized in graphs:
      http://ci5.iarc.fr/CI5plus/old/Graph4p.asp?cancer%5B%5D=100&male=1&female=2&country%5B%5D=5280099&sYear=1950&eYear=2010&stat=3&age_from=4&age_to=18&orientation=1&window=1&grid=1&line=2&moving=1&scale=0&submit=%C2%A0%C2%A0%C2%A0Execute%C2%A0%C2%A0%C2%A0

    o Head and Neck Cancer incidence, prevalence, mortality and risk from year 1990 through 2015 can be retrieved from the Netherlands Cancer Registry web site, which is managed by the Netherlands Comprehensive Cancer Organisation (IKNL). The registration of data in this database is performed by registration employees of IKNL http://www.cijfersoverkanker.nl/?language=en. Data can be exported in tables and visualized in graphs. The query can be customized but only the number of cases can be retrieved online (no additional data). It can be used for aggregated data on population. No risk factors or lifestyle behaviours are part of the selection criteria. Clinicians, physician epidemiologists or policymakers can request customized data from the Netherlands Cancer Registry (NKR). The data can also be linked to patient data. The data are provided in accordance with Dutch privacy regulations. Each request is checked against the policy IKNL applies for use of data from the NKR.

- Slovenia

    o Oral cavity and pharynx:
      http://ci5.iarc.fr/CI5plus/old/Graph4p.asp?cancer%5B%5D=20&male=1&female=2&country%5B%5D=7050099&sYear=1950&eYear=2010&stat=3&age_from=4&age_to=18&orientation=1&window=1&grid=1&line=2&moving=1&scale=0&submit=%C2%A0%C2%A0%C2%A0Execute%C2%A0%C2%A0%C2%A0

    o Larynx:
      http://ci5.iarc.fr/CI5plus/old/Graph4p.asp?cancer%5B%5D=100&male=1&female=2&country%5B%5D=7050099&sYear=1950&eYear=2010&stat=3&age_from=4&age_to=18&orientation=1&window=1&grid=1&line=2&moving=1&scale=0&submit=%C2%A0%C2%A0%C2%A0Execute%C2%A0%C2%A0%C2%A0

- Germany

    o Data from German Cancer Registries can be extracted and visualized from the German Centre for Cancer Registry Data (ZfKD: http://www.krebsdaten.de/Krebs/SiteGlobals/Forms/Datenbankabfrage/EN/datenbankabfrage_stufe2_form.html. Data can be exported in tables and visualized in graphs. The query can be customized but only the number of cases can be retrieved online (no additional data). It can be used for aggregated data on population. No risk factors

or lifestyle behaviors are part of the selection criteria.

Scope of data

From these data in BD2Decide we might analyze the general correlation of deaths with other risk factors (e.g. alcohol consumption, smoking, exposure to pollutants) and check how population data correlate with high/resolution data from BD2Decide study cohorts. Also, by combining this population aggregated information with CRs and with BD2Decide patients' data our statisticians and Big Data analysts might identify how population habits or geographic conditions influence Head and Neck cancer onset and prognosis.

Data relevant to BD2Decide Countries must be selected.

3.1.1　Genomics and imaging public datasets

Image diagnostics datasets are publicly available at the following sources:

- TCIA-The Cancer Imaging Archive, http://www.cancerimagingarchive.net/ : forthcoming dataset

- MAASTRO images database (https://www.cancerdata.org/data?q=image_archive), that contains images of the Head and Neck region, usable for radiomics training.

These datasets should be used to validate or to refine the data analysis algorithms.

Gene sequencing datasets where Head and Neck cancer cases are available refer to the most famous gene sequencing project: The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/tcga/tcgaCancerDetails.jsp?diseaseType=HNSC&diseaseName=Head and Neck squamous cell carcinoma). Here we can extract genomic data to be used jointly or comparatively with the genomic data extracted from BD2Decide cohorts´ biological specimens by INT.

3.1.2　Lifestyle and risk factors

EUROSTAT provides some aggregated data on health determinants that might be of use to correlate populations lifestyle behaviors with disease prognosis.

Data can be downloaded from EUROSTAT datasets at URL: http://ec.europa.eu/eurostat/statistics-explained/index.php/Health by selecting for example Tobacco smoking habits table (online data code: hlth_ehis_de3).

Country(	Total	Men	Women	Gender difference
		(%)		(percentage points)
Belgium (¹)	18,9	21,1	17,0	4,1
Bulgaria	29,2	40,4	18,9	21,5
Czech Republic	24,3	29,6	19,4	10,2
Denmark	:	:	:	:
Germany	22,8	25,5	20,3	5,2
Estonia	25,9	39,5	15,1	24,4
Ireland	:	:	:	:

Country(	Total	Men (%)	Women	Gender difference (percentage points)
Greece	31,8	37,8	26,1	11,7
Spain	25,2	29,5	21,0	8,5
France	:	:	:	:
Croatia	:	:	:	:
Italy	:	:	:	:
Cyprus	25,9	37,9	14,3	23,6
Latvia	27,9	46,0	13,0	33,0
Lithuania	:	:	:	:
Luxembourg	:	:	:	:
Hungary	26,1	31,4	21,5	9,9
Malta	19,2	23,8	15,1	8,7
Netherlands	:	:	:	:
Austria	22,9	26,8	19,3	7,5
Poland	23,8	30,9	17,9	13,0
Portugal	:	:	:	:
Romania	20,5	32,7	9,1	23,6
Slovenia	18,7	22,1	15,5	6,6
Slovakia	19,3	26,9	12,3	14,6

Table 3-2. Example: Proportion of daily smokers of cigarettes, persons aged 15 and over, by sex and age, 2008 or nearest year (%). Source: EUROSTAT

([1]) Data with low reliability.

These set of tables also considers smoking as correlated to education level. This combined information might be linked to BD2Decide patients´ cohorts data to identify possible correlations with disease outcomes or other endpoints (e.g. QoL).

These tables have been downloaded and can be used by BD2Decide statisticians.

3.1.3 Environmental factors

The main sources of data are:

- Italy: environmental observatory at ISPRA (Italy)

- Regional Environment Monitoring Agencies in Italy (ARPA, Regional Governments - Italy)

- EUROSTAT: levels of exposure to pollutants by Country

- WHO statistics concerning population exposures to some pollutants

- YALE University environmental performance index.

The environmental observatory at ISPRA (Italy) provides a number of data concerning concentration of pollutants in water and in air by Region and on the entire Italian territory. These resources, combined with the residence of BD2Decide cohorts, might increase the prognostic effectiveness of models.

Different data are available on ISPRA databases, among which we are interested in the following, as already indicated in deliverable D2.1:

- Databank of air pollutants from traffic- Italy. Source ISPRA. Consultation by pollutant: http://www.sinanet.isprambiente.it/it/sia-ispra/fetransp/

- Pollutants emissions years 1990-2014 (Excel file): http://www.sinanet.isprambiente.it/it/sia-ispra/serie-storiche-emissioni/serie-storiche-delle-emissioni-nazionali-di-inquinanti-atmosferici/view

- Air pollutants (SNAP series) year 1980-2014 - Italy (excel file): http://www.sinanet.isprambiente.it/it/sia-ispra/serie-storiche-emissioni/serie-storiche-delle-emissioni-nazionali-snap/view

- Air pollution data by region, pollutant and year can be found for Italy at URL: http://www.brace.sinanet.apat.it/web/struttura.html?p_livello_1=18&p_main=web/area_download.inizio&p_scroll=. The database must be queried and data retrieved. To use these datasets in BD2Decide, they should be all downloaded unless methods to query them and extract data are available.

- ARPA Emilia Romagna (Excel): pollution data in Emilia/Romagna region http://www.arpa.emr.it/dettaglio_generale.asp?id=2615&idlivello=1521 and also https://www.arpae.it/qualita-aria/bollettino-qa-/?idlivello=1924 for detailed pollution data. This web DB can be queried online. Data can be also exported for data analysis.

- ARPA Lombardia (Excel) provides day/by/day information on pollution and emissions by pollutant and origin of emission, downloadable from URL: http://shp.arpalombardia.it/sites/arpalombardia2013/RSA/Pagine/default.aspx. Data at district level are available for query or download from URL: http://inemar.arpalombardia.it/inemar/webdata/elab_standard_prov.seam;jsessionid=0D30D6FCB014B0DDA9E98434812CE232?cid=175137

- Regione Lombardia provides also other aggreagated data (CSV, PDF, XLSX, XML) at URL: https://www.dati.lombardia.it/browse?category=Ambiente.

| DATO: EMISSIONI INQUINANTI PER MACROSETTORE | FONTE DATI: INEMAR ARPA Lombardia | ANNO DI RIFERIMENTO: 2012 (dati per revisione pubblica) | | | | | | | | | | | | | |

MACROSETTORE	PROV	SO₂ t/anno	NOₓ t/anno	COV t/anno	CH₄ t/anno	CO t/anno	CO₂ t/anno	N₂O t/anno	NH₃ t/anno	PM2,5 t/anno	PM10 t/anno	PTS t/anno	SOST ACIDIFICANTI kt(H+)/anno	PREC. O₃ t/anno	GAS SERRA kt(CO₂ eq)/anno
Produzione di energia e trasform. combustibili	BG	2	128	14	102	91	35	4		2	2	2	3	181	39
Combustione non industriale	BG	101	1.456	1.683	1.211	13.904	1.739	82	35	1.543	1.567	1.648	37	5.005	1.790
Combustione nell'industria	BG	1.012	3.457	335	97	1.413	1.775	106	111	81	102	140	113	4.708	1.810
Processi produttivi	BG	518	226	717	18	8.345	1.139	5	0	87	184	231	21	1.910	1.141
Estrazione e distribuzione combuastibili	BG			696	8.180									811	172
Uso di solventi	BG	0	15	9.273	0	8			0	87	98	150	0	9.291	125
Trasporto su strada	BG	11	6.545	1.715	139	7.795	1.817	64	111	384	510	649	149	10.559	1.840
Alttre sorgenti mobili e macchinari	BG	28	1.153	121	2	530	148	5	0	44	45	45	26	1.586	149
tTrattamento e smaltimento rifiuti	BG	174	610	16	6.998	227	198	56	31	6	6	8	20	883	362
Agricoltura	BG	0	28	4.144	16.388	0		901	8.184	29	73	144	482	4.408	624
Altre sogrenti e assorbimenti	BG	3	13	6.240	236	340	-784	0	2	117	120	122	0	6.297	-779
Produzione di energia e trasform. combustibili	BS	201	422	33	195	214	265	12	1	23	28	29	16	574	273
Combustione non industriale	BS	116	1.645	1.571	1.097	11.777	2.087	84	31	1.284	1.306	1.373	41	4.888	2.136
Combustione nell'industria	BS	2.238	3.829	581	103	4.263	2.252	80	20	99	127	181	154	5.722	2.279
Processi produttivi	BS	703	1.126	1.773	91	11.879	941	31	5	147	314	374	47	4.455	952
Estrazione e distribuzione combuastibili	BS			889	9.315									1.019	196
Uso di solventi	BS	1	3	9.129		3			0	84	89	135	0	9.134	142
Trasporto su strada	BS	17	9.877	2.161	180	10.805	2.719	88	184	566	754	970	226	15.402	2.750
Alttre sorgenti mobili e macchinari	BS	12	2.387	239	4	757	211	12	1	120	123	126	52	3.234	214
tTrattamento e smaltimento rifiuti	BS	101	478	572	32.175	168	235	32	34	4	4	5	16	1.624	920
Agricoltura	BS		68	12.994	53.613			2.815	27.212	105	266	530	1.602	13.827	1.998
Altre sogrenti e assorbimenti	BS	3	17	11.636	1.960	440	-1.207	0	3	141	145	148	1	11.732	-1.166

Table 3-3. Example of air pollution emissions by district, Lombardy region, year 2012

Besides the Italian situation, which will be analyzed in details by geographic area as combined to individual patients´ data of BD2Decide cohorts, other databases are available where data concerning pollution in EU Countries can be found (aggregated indicators by Country & pollutant & year):

- Environmental performance index – Netherland, Italy (Excel file): http://epi.yale.edu/country/netherlands and http://epi.yale.edu/country/netherlands

- Eurostat: levels of exposure to pollutants by Country are available at URL: http://ec.europa.eu/eurostat/tgm/table.do?tab=table&plugin=1&language=en&pcode=tsdph3 70. The source dataset can be queried to extract any combination of available data (http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=env_air_emis&lang=en).

airpol | Particulates < 2.5µm ▾

geo \ time	2007	2008	2009	2010	2011	2012	2013	2014
EU (28 countries)	16.7	17.5	17.4	18	18.4	16.8	15.9	15.2
Belgium	21.1	18.6	18.9	17.7	17.7	16.1	16.8	14.2
Bulgaria	41.2	41.5	24.7	31.1	41.3	29.3	26.5	26.1
Czech Republic	18	19.3	20.3	22.8	21	19.2	20.4	19
Denmark	10.3	10.8	10.2	11	16.3	11.1	10.1	12.4
Germany	15.3	16.1	16.7	17.4	17.1	14.2	14.6	15.1
Estonia	10.9	5.4	6.2	7.6	6.9	7.8	8.1	8.6
Ireland	:	:	9.6	10.9	9	:	10.3	7.8
Greece	26.2	24.3	20	:	17	18.4	:	:
Spain	13.1	14.2	15.2	12.4	12.8	13.5	10	11.1
France	12.9	15.5	18.1	18.3	17.8	16	15.4	12.6
Croatia	:	:	:	:	:	:	:	:
Italy	28.8	25.1	25.1	23.3	26.8	23	20.1	17.5

Table 3-4. Example: Urban population exposure to air pollution by particulate matter Code: tsdph370 (Micrograms per cubic metre). Source EUROSTAT

- WHO statistics are also available that measure the % of population exposed to air pollutants such as PM10 or PM2.5 by Country and sex (year series available). These data can be accessed and downloaded in excel format from the public website of the Environment and Health Information System (ENHIS): http://data.euro.who.int/eceh-enhis/Default2.aspx?indicator_id=18&subindicator_id=40&type=Bar&period=2004&NUTS=1.

  o Exposure to particulate matter in outdoor air-

  o Exposure to chemical hazards in food-Heavy metal intake in adults (see Figure 3-1 below).
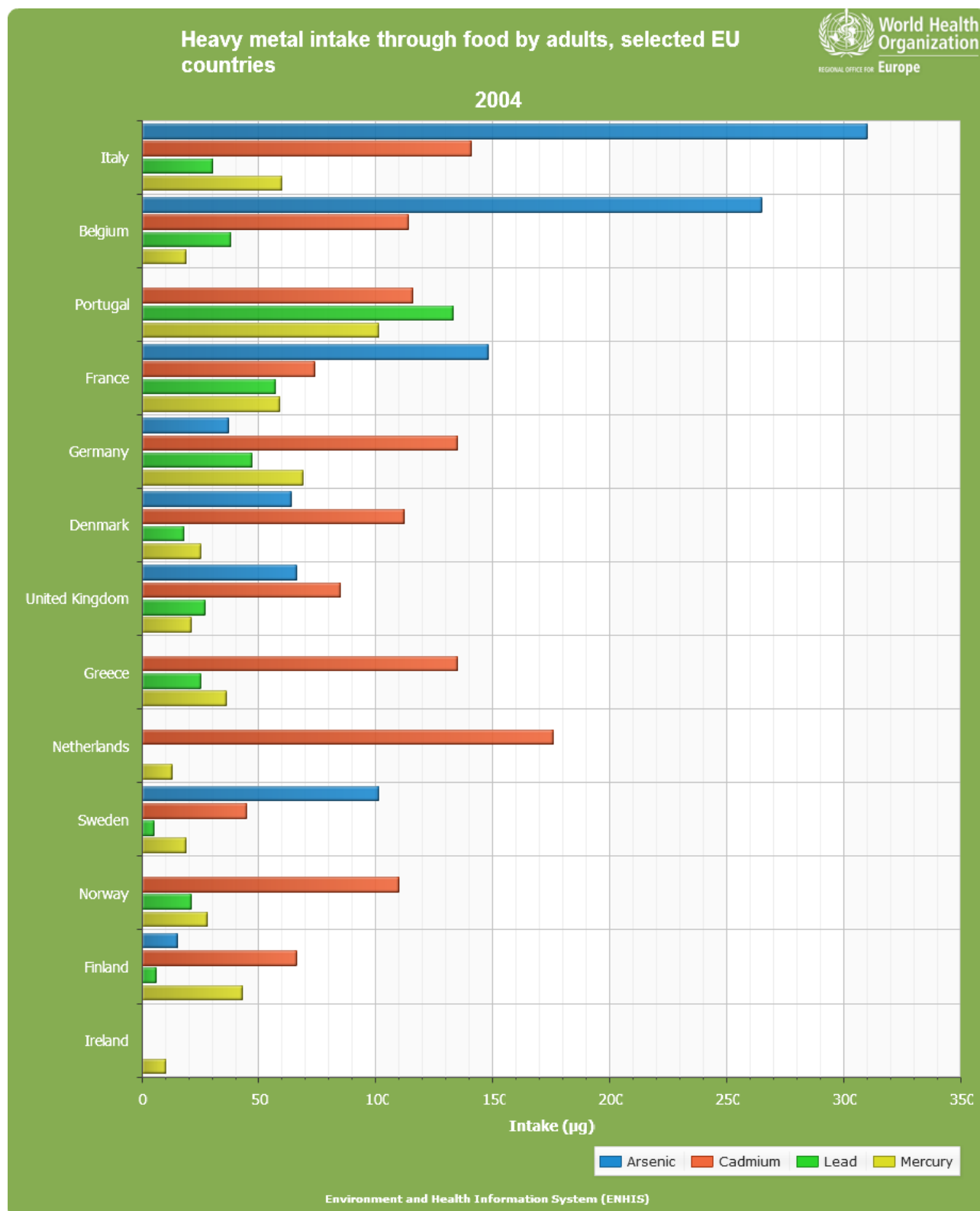
Figure 3-1. Graph depicting the % of population with metal intake in foods by Country (year 2004, source WHO)

We have downloaded and made available to BD2Decide the majority of above mentioned datasets for use in BD2Decide.

3.1.4 Public datasets useful for economic impacts (DALYs)

The WHO gives access to a DB with the latest global, regional and country-level estimates of cause-specific disability-adjusted life year (DALYs), years of life lost (YLL) and years lost due to disability (YLD) for the year 2000 and 2012. These data might be used by VUmc to tailor the estimates of burden of disease. Data can be downloaded from the URL: http://apps.who.int/healthinfo/global_burden_disease/estimates/en/index2.html. Data are aggregated by WHO Regions and disease, therefore are not very precise but might provide a rough estimate of disease burden.

In this case we should consider the table: Global Health Estimates (GHE) 2014: DALYs by age, sex and cause only the rows relevant to Malignant neoplasms and in particular "Mouth and oropharynx cancers" that provides aggregated data by sex and age.

We have downloaded these tables for data analysis. Data for populations relevant to BD2Decide study must be selected. This information might be used to estimate the potential average socio/economic impacts of BD2Decide improved outcome prediction (e.g. reduced post/treatment morbidity or avoided treatment consequences, reduced hospital stay etc.).

Both sexes Total		Age group	0-27 days	1-59 months	5-14 years	15-29 years	30-49 years	50-59 years	60-69 years	70+ years	Total
Malignant neoplasms	**48.398.025**	**Male**	**31**	**158.646**	**186.712**	**478.607**	**2.618.231**	**6.562.829**	**8.203.226**	**9.388.377**	**27.596.659**
Mouth and oropharynx cancers	1.567.013		0	793	2.464	11.049	189.702	487.134	381.580	193.132	1.265.854
Malignant neoplasms		**Female**	**45**	**124.157**	**145.834**	**379.921**	**2.848.666**	**4.569.747**	**5.134.228**	**7.598.768**	**20.801.366**
Mouth and oropharynx cancers	0		1.280	1.210	8.863	47.612	83.812	74.937	83.445	301.159	0

Table 3-5. DALY estimates, 2000–2012 by WHO Region

The same information (estimated DALYs), is also available by Country, sex and age (table below, GHE_DALY_2012_country) at URL http://apps.who.int/healthinfo/global_burden_disease/estimates/en/index2.html.

World Health Organization
Department of Health Statistics and Information Systems
May 2014

Estimated DALYs ('000) by cause, sex and WHO Member State (1), 2012

| Sex	GHE code	GHE cause	Member State (See Notes for explanation of colour codes)	Germany	Ghana	Greece	Guatemala	Guinea	Guinea-Bissau	Guyana	Haiti	Honduras	Hungary	Iceland
			WHO Country code	4085	1180	4140	2250	1190	1192	2260	2270	2280	4150	4160
Persons	61	A. Malignant neoplasms		5017,4	373,3	596,9	302,6	136,9	22,5	18,8	180,1	156,5	833,0	12,9
Persons	62	1. Mouth and oropharynx cancers		168,9	10,1	7,6	4,6	3,2	0,5	0,4	3,4	1,7	49,3	0,2
Persons	63	2. Oesophagus cancer		141,2	2,7	4,7	3,5	0,6	0,2	0,3	3,7	1,5	18,6	0,4
Persons	64	3. Stomach cancer		227,7	15,2	26,6	56,4	5,8	1,0	0,7	15,0	21,9	39,3	0,4
Persons	65	4. Colon and rectum cancers		566,4	13,2	46,3	8,0	2,5	1,0	1,0	9,1	6,7	118,6	1,3

Table 3-6. WHO estimated of DALYs by WHO member State (male and female data by sex)

4 MANAGEMENT OF THE EXTERNAL DATA SOURCES IN BD2DECIDE

This chapter describes how the external data are treated and managed within the BD2Decide system. By external we mean information that is not collected within hospital settings and related to the H&NC patients involved in the retrospective and prospective studies, as shown in **Errore. L'origine riferimento non è stata trovata.**.

More information about how the external data sources are going to be integrated and used by the different modules of the system, in both the service and user interfaces layers, is provided in D2.3.
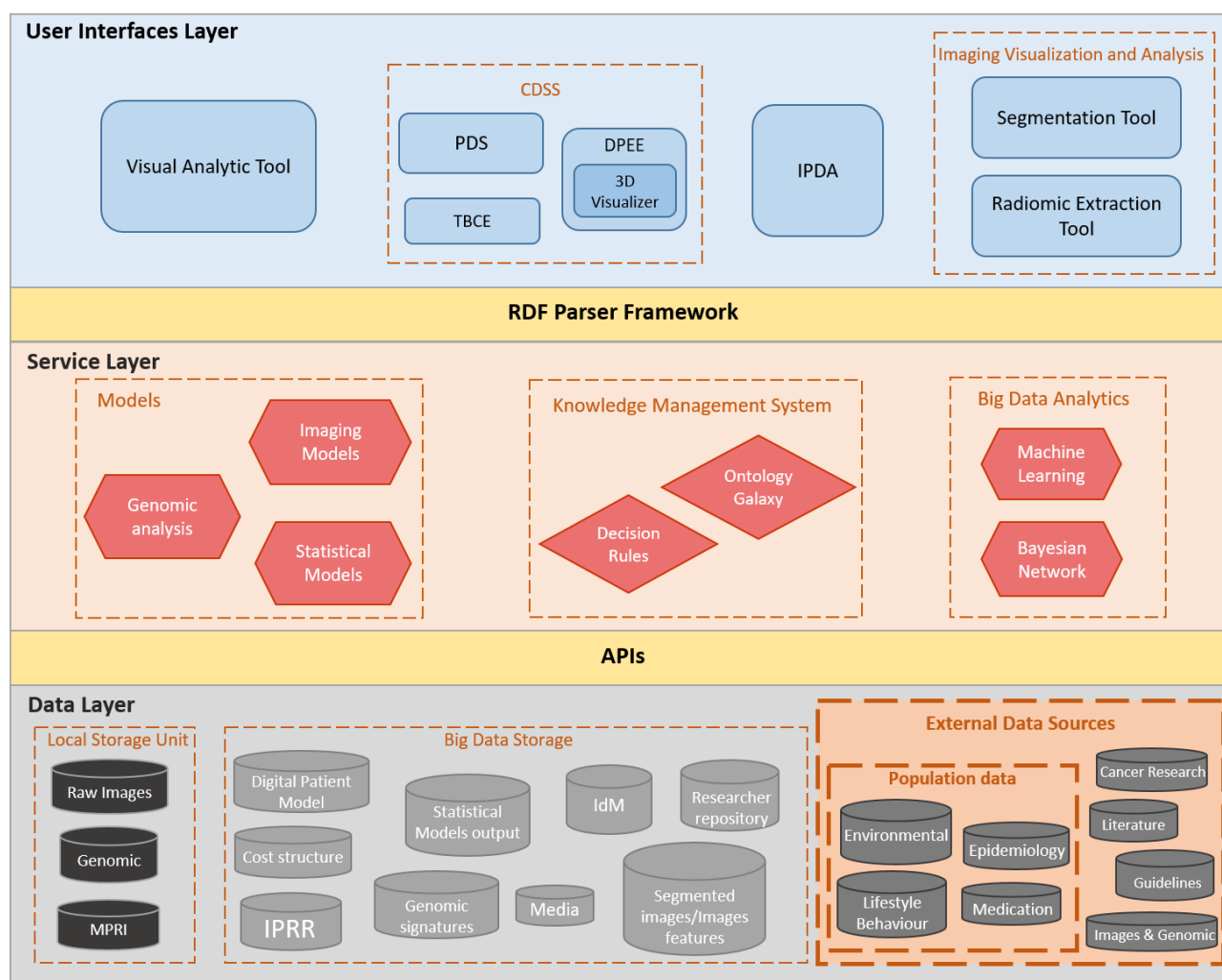
Figure 4-1- High level overview of the External Data Sources (highlighted in the bottom right side of the figure) in the BD2D system, taken from D2.3

4.1 CANCER REGISTRIES (UPM/VUMC/AII)

4.1.1 INT Cancer Registry

INT Cancer Registry data is stored in a cloud repository accessible via ftp for research and clinical activities. The information provided can be grouped in two categories, as defined before: *Pilot study* and *High-resolution study*. Both groups of data feed the BD2Decide system, after anonymization, in different ways. As data containing epidemiology information of individual patients, *Pilot study* data

is used to test the BD2Decide statistical models (developed by VU/Vumc partner) and to train the Big Data Analytics (BDA) algorithms. On the other hand, *High-resolution study* is used to feed VU/Vumc models, while the BDA will employ them within the decision support system (both in Clinical Decision Support System and Visual Analytics Tool).

4.1.2 External Cancer Registries

In the context of BD2Decide, this data can be useful for improving the statistical outcomes of the models, for the BDA processes or to complement the BD2Decide dataset. External CR usage can enhance the queries outcomes, being the more information the better refinement for the decision support system outcomes.

Similarly to INT CR data, external CR data are collected and stored in a cloud repository, accessible via ftp for research and clinical activities. Data are also used to feed the statistical models and to be processed by the BDA for CDSS and the VAT. Additionally, external CR will be employed to display Head and Neck Cancer (H&NC) context overview, like the percentage of H&NC patients group by country.

4.1.3 External Images and Genomic databases

The external imaging datasets TCIA and MAASTRO are accessed by the CDSS and in the VAT as following:

1. Access via website link, through the integration of datasets APIs into the BD2Decide tools. This solution allows to work directly with the data, without integrating it in the system. Then, data can be downloaded in a local PC to be analysed by the BD2Decide tools or outside the BD2Decide tools.

2. Collect the imaging data in a cloud repository for research and/or clinical activities.

In case the data is stored in a cloud repository, it is possible to access the raw images through a tool integrated in the CDSS and the VAT.

The possibility of aggregate and analyse external image datasets of the segmented and radiomics features allow to enrich the analysis outcomes.

4.2 ENVIRONMENTAL, LIFE BEHAVIOUR AND MEDICATION DATA

4.2.1 ISS Data

Population data, collected by ISS, is grouped in two types of data, as mentioned previously: *Aggregated data* and *High-resolution data*. Both groups are used to refine the statistical models and by the BDA with the aim of obtaining outcomes that may help clinicians on treatment decision.

Aggregated data contains general health status information, used to display H&NC main context overview, like the average of patients per region, and the number of new H&NC incidences per year.

As for the INT CR data, ISS data is collected in a scientific cloud, accessible via ftp for research and analysis purpose.

4.2.2 External dataset

ISPRA, ARPA, EUROSTAT, WHO and Yale University environmental performance index are public datasets containing environmental, life behavior, risk factors and socioeconomic data. These data will be used in the models to improve the statistical outcomes and can be processed by BDA algorithms to identify cancer correlation together with the BD2Decide data.

As for the ISS data, these external datasets are collected in a cloud, accessible via ftp for research and clinical purpose. Once these data are stored in the BD2Decide infrastructure, statistical models and BDA can access them. It is also possible to access the data to provide H&NC status overview in the VAT, for instance: percentage of H&NC patients by country and/or by year or contrasting the number of H&NC patients with the pollution.

4.3 DATA FROM LITERATURE

Scientific literature is key to widen the researchers' knowledge and to allow them being up to date on their activities and studies. This aspect justify how important is the integration of literature scientific contents in the VAT.

The employment of literature materials in the VAT is represented in two ways:

1. New scientific papers related to the researcher work.

2. A scientific literature research engine.

With the purpose of achieving the first point, a pre-set of BD2Decide keywords is available. The BD2Decide scientific keyword is intended to aid the researcher with a pattern of the index entries to be used for his/her research. The definition of the BD2Decide keywords that is currently in progress, allows the researcher to recommend and collect new scientific literature. Also, researcher can customize a keyword, by editing, deleting or adding it in order to personalize the research activities.

The literature search is carried out by accessing to an external literature search engine, like PubMed or EHNS, through specific Application Programming Interfaces (APIs). The external APIs are integrated in the platform by All In Image partner, as responsible for the Big Data Infrastructure, and allow the researcher to look for specific literature, specifying the publication research of interest. Apart from producing a list of literature recommended, a pre-visualization of the manuscripts found are provided, as a result of the Big Data Analytics algorithms. This pre-visualization will include scientific information useful for the researcher, such as how many publications include a certain keyword. An analysis of the data included in the literature, such as statistical results of applying specific treatments to a group of patients, is also included. The pre-visualization will be carried out through content extraction and analysis solutions performed by All In Image partner and will help the researcher to give a preliminary description of the search, before reading the literature results.

4.4 CLINICAL GUIDELINES

Clinical guidelines aim at helping clinicians to make the best decisions about treatment and care for patients. In H&NC, like in other types of cancer, it is usual to consult clinical guidelines for the treatment and the care decision's assistance. The following guidelines have been chosen by clinical partners:

- American Society of Clinical Oncology (ASCO)

- The European Society for Medical Oncology (ESMO)

- The European Head and Neck Society (EHNS)

- European SocieTy for Radiotherapy & Oncology (ESTRO)

- The National Comprehensive Cancer Network (NCCN)

Guidelines can be read through the CDSS and periodically reviewed and updated, by the corresponding organization, following one of the options:

1. Clinicians with the proper rights (only a clinician or a group of clinicians will be able to perform this action) should upload the latest clinical guidelines to the CDSS, where it will be available for the rest of clinicians.

2. Once the latest version of the clinical guidelines are updated, an automatic process will extract and update them in the CDSS.

Furthermore, a set of basic rules are going to be extracted from the guidelines, with the aim of feeding the Knowledge Management System and then providing this additional information to the decision support system, in the CDSS and in the VAT.

5   ETHICS

We will mostly use aggregated population data, not linked to individuals, therefore ethics aspects have been considered only for the cases of high-resolution studies outlined in sections 1.1 and 1.2 above. The main ethical aspects concern the safeguard of privacy and patients' dignity and the respect of patient's consent to access personal data for research purposes.

Patients involved in high-resolution studies, for the cases where this is applicable, will provide written consent to access their data for research and statistical analysis. In most cases such consent has already been provided to the data custodians (i.e. the institutions that collect and manage the data), thus INT and ISS will request permission to access data for epidemiology studies and for BD2Decide data analysis directly to the data custodians.

5.1   ETHICS ASPECTS RELATED TO CANCER REGISTRIES DATA

Cancer Registries are maintained by National or Regional Institutions established in each European Country. Not all European Territory is covered, although the coverage is growing.

Duet to this fragmentation in cancer registries data collection and management, and due to different legal and ethics regulatory frameworks across Europe, access to cancer registries data must be requested to the managing institutions.

Cancer registries data are collected for each individual and thus include a unique identifier (e.g. NHS insurance code or tax code) used to cross-check records whenever possible.

The cancer registries generally do not supply data by name. They do supply, for well-justified requests from specific institutions and national and international publications, aggregated data without identifying information. Subject to the consideration of the registry director and approval by the ethics committee, the registry can supply identifying information to specified researchers for particular research purposes described in detail in a project that lays out the specific aims of the research.

INT has requested access to the CRs mentioned in this document and has provided to the relevant CR management institutions the BD2Decide study protocol and a detailed explanation of the use of such data, of the privacy and pseudonymization procedures implemented by the project and of the data management in BD2Decide datasets. INT was also requested to communicate the names of the partners who will hold the CR data, the location where data are stored.

For high-resolution studies, INT has already achieved permissions for most of the CRs contributing to the RARECAREnet studies.

Data - if present - concerning ethnicity, sexual habits, religion or any other sensitive information whose disclosure is not allowed by EU and/or National laws, will not be extracted for BD2Decide data analysis.

The link between the patient identity and the collected data will be maintained by the institutions having access to such data and will not be available to other BD2Decide partners.

5.2 ETHICS ASPECTS RELATED TO COLLECTION OF PATIENT'S HEALTH-RELATED DATA FROM NHS

Patients' health related data will be collected for the high-resolution study conducted by ISS on a small cohort of selected Italian patients as described above (see section 2).

ISS has established contacts with some Italian Territory Health Agencies (AUSL), which are the custodians of all patients' health data generated from GPs and specialized physicians, pharmacies and diagnostic laboratories, in order to acquire data for a selected cohort of Italian patients from INT and AOP.

To collect high-resolution data, i.e. data linked to individual patients, the tax code and the NHS insurance number will be used to cross-check patient's data integration. Once collected, data will be pseudonymized and assigned the BD2Decide ID code that will not allow to disclose the patient's identity.

The link between the patient identity and the collected data will be maintained by the institutions having access to such data and will not be available to other BD2Decide partners.

6 REFERENCES

[1] Forsea A-M, Cancer registries in Europe—going forward is the only option. Ecancermedicalscience. 2016; 10: 641. www.ncbi.nlm.nih.gov/pmc/articles/PMC4898937/

[2] Dimitrova N., European Network of Cancer Registries, Steering Committee, Population based cancer registries in Europe.

[3] Gatta G, Botta L, Sánchez MJ, Anderson LA, Pierannunzio D, Licitra L; EUROCARE Working Group. Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: The EUROCARE-5 population-based study. Eur J Cancer. 2015 Sep 6. pii: S0959-8049(15)00749-2. doi: 10.1016/j.ejca.2015.07.043.

[4] Van Dijk BA, Gatta G, Capocaccia R, Pierannunzio D, Strojan P, Licitra L; RARECARE Working Group. Rare cancers of the head and neck area in Europe. Eur J Cancer. 2012 Apr;48(6):783-96. doi: 10.1016/j.ejca.2011.08.021. Epub 2011 Nov 1.

[5] Al-Dakkak I. Socioeconomic status and head and neck cancer. Evid Based Dent. 2010;11(2):57-8. doi: 10.1038/sj.ebd.6400726.

[6] Choi SH, Terrell JE, Fowler KE, McLean SA, Ghanem T, et al. (2016) Socioeconomic and Other Demographic Disparities Predicting Survival among Head and Neck Cancer Patients. PLOS ONE 11(3): e0149886. doi: 10.1371/journal.pone.0149886

[7] Conway DI and al., Socioeconomic factors associated with risk of upper aerodigestive tract cancer in Europe. Eur J Cancer. 2010 Feb;46(3):588-98. doi: 10.1016/j.ejca.2009.09.028. Epub 2009 Oct 24.

[8] Báez A. Genetic and environmental factors in head and neck cancer genesis. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 2008 Apr-Jun;26(2):174-200

[9] Churilla, T.M. et al., The Impact of Health Insurance Status on the Presentation, Local Management, and Outcomes of Patients With Head and Neck Cancer in the United States. International Journal of Radiation Oncology • Biology • Physics , Volume 94 , Issue 4 , 870

[10] Zhan, K. et al., Demographics, Disparities, and Survival in Young Patients With Oral Cavity Squamous Cell Carcinoma: A Population-Level Analysis of 3828 Cases. International Journal of Radiation Oncology • Biology • Physics , Volume 94 , Issue 4 , 912 - 913

[11] Langevin SM, McClean MD, Michaud DS, Eliot M, Nelson HH, Kelsey KT. Occupational dust exposure and head and neck squamous cell carcinoma risk in a population-based case–control study conducted in the greater Boston area. Cancer Medicine. 2013;2(6):978-986. doi:10.1002/cam4.155.

[12] Sapkota A. et al., Indoor air pollution from solid fuels and risk of upper aerodigestive tract cancers in Central and Eastern Europe. Environmental Research 120 (2013) 90-95.

[13] Wong ICK, Ng Y-K, Lui VWY. Cancers of the lung, head and neck on the rise: perspectives on the genotoxicity of air pollution. Chinese Journal of Cancer. 2014;33(10):476-480. doi:10.5732/cjc.014.10093.

ANNEXES

CANCER REGISTRY DATA FOR HIGH RESOLUTION STUDY

country	regis	id	gender	dateofbirth	dateofdiagnose	primarytumoursiteode4di
Italy	Parma	113379	1	02/06/1953	12/01/2009	3:
Italy	Parma	113389	1	01/11/1958	14/01/2009	
Italy	Parma	113406	1	16/08/1945	30/01/2009	3:
Italy	Parma	113484	1	19/09/1931	18/03/2009	
Italy	Parma	113586	0	31/01/1961	27/03/2009	
Italy	Parma	113593	1	17/10/1953	07/04/2009	3:
Italy	Parma	113602	1	30/04/1940	29/05/2009	
Italy	Parma	113725	1	06/06/1956	11/06/2009	
Italy	Parma	113809	1	25/04/1960	03/06/2009	
Italy	Parma	113813	1	06/09/1940	19/10/2009	
Italy	Parma	114124	1	18/03/1939	01/02/2009	3:
Italy	Parma	114150	0	11/09/1935	02/04/2009	
Italy	Parma	114214	1	11/02/1943	15/09/2009	3:
Italy	Parma	114215	1	30/03/1930	20/10/2009	3:
Italy	Parma	114220	1	27/06/1931	15/12/2009	3:
Italy	Parma	115365	0	19/12/1931	20/04/2009	
Italy	Parma	115367	1	19/10/1943	27/10/2009	
Italy	Parma	115369	1	05/05/1949	14/09/2009	
Italy	Parma	115374	0	11/12/2020	22/04/2009	
Italy	Parma	115376	0	25/04/2024	16/06/2009	
Italy	Parma	115378	1	24/01/1954	15/10/2009	
Italy	Parma	115379	0	21/07/1944	10/11/2009	
Italy	Parma	115382	1	21/10/1937	26/05/2009	
Italy	Parma	115383	0	04/09/1941	17/04/2009	
Italy	Parma	115389	1	28/12/1950	01/04/2009	

| DB | codebook | ns removed | problems with the date | (+) |

Figure Annex-1. Codebook for high resolution study (Parma cohort)

DB available includes also the additional variables defined for the high resolution study analyses.

Site	Stage	Treatment
larynx	advanced	combined treatments beyond 3 months
oral_cavity	advanced	surgery+radio or + concomitant radio-chemotherapy
larynx	Metastatic	only surgery
oropharynx	advanced	surgery+radio or + concomitant radio-chemotherapy
oral_cavity	advanced	surgery+radio or + concomitant radio-chemotherapy
larynx	advanced	combined treatments beyond 3 months
oral_cavity	advanced	combined treatments beyond 3 months
oral_cavity	advanced	combined treatments beyond 3 months
oropharynx	Metastatic	only surgery
oropharynx	Metastatic	surgery+chemotherapy within 2 months
larynx	other	only surgery

Table Annex-1. Data available for high-resolution study

DESCRIPTION OF AGGREGATED POPULATION DATA PROVIDED BY ISS (ITALY)

1. Demographic

Resident population:

There are two indicators of the resident population, calculated as follows:

(a) $P_{resident} = (Pop_{01/01/t} + Pop_{01/01/t+1})/2$

It is the arithmetical average of the populations at first January of the year t and at first January of the year t+1.

Stratification variables are: gender (males, females), age group (0-4, 5-14, …, 75+)

Available in: 1982-2015 (annual statistics),

At geographical level: province and region.

Data Source:

- period 1982-1991: ISTAT, population is reconstructed using census data 1981, 1991 and anagraphical records

- period 1992-2001: ISTAT population is reconstructed using census data 1991, 2001 and anagraphical records

- from 2002 on: ISTAT, using the administrative survey entitled "Rilevazione della popolazione residente comunale per sesso, anno di nascita e stato civile al 31/12/XX ricostruita secondo i dati del Censimento della popolazione del 2001 e stato civile al 31 dicembre"

Note that the indicator is referred to the geographical area of residence.

(b)   % distribution by age of Population resident: %Population resident,x = [President, x]/[President]*100

It represents the percent distribution of the resident population of age x at year t, calculated as ratio of resident population of age x in year t and total resident population in year t.

Stratification variables are: gender (males, females), age group (65+, 85+)

Available in: 1982-2015 (annual statistics),

At geographical level: province and region

Data Source:

- period 1982-1991 : ISTAT, population is reconstructed using census data 1981, 1991 and anagraphical records

- period 1992-2001: ISTAT population is reconstructed using census data 1991, 2001 and anagraphical records

- from 2002 on: ISTAT, using the administrative survey entitled "Rilevazione della popolazione residente comunale per sesso, anno di nascita e stato civile al 31/12/XX ricostruita secondo i dati del Censimento della popolazione del 2001 e stato civile al 31 dicembre"

Note that the indicator is referred to the geographical area of residence.

Dependency ratio

There are three indicators of dependency calculated for the year t as follows:

$DR_1 = (Pop_{0-14} + Pop_{65+})/(Pop_{15-64}) * 100$

$DR_2 = (Pop_{65+})/(Pop_{15-64}) * 100$

$DR_3 = (Pop_{0-14})/(Pop_{15-64}) * 100$

Stratification variable is gender (males, females)

Available in: 1982-2015 (annual statistics),

At geographical level: province and region

Source:

- period 1982-1991: ISTAT, population is reconstructed using census data 1981, 1991 and anagraphical records

- Years 1992-2001: ISTAT population is reconstructed using census data 1991, 2001 and anagraphical records

- from 2002 on: ISTAT, using the administrative survey entitled "Rilevazione della popolazione residente comunale per sesso, anno di nascita e stato civile al 31 dicembre", based on census data 2001

Note that the indicator is referred to the geographical area of residence.

Ageing rate

The Ageing rate (AGrate) is calculated as follows:

$\mathbf{AGrate} = (Pop_{65+})/(Pop_{0-14}) * 100$

Stratification variable is gender (males, females)

Available for the period: 1982-2015 (annual statistics)

At geographical level: province and region

Source:

- period 1982-1991: ISTAT, population is reconstructed using census data 1981, 1991 and anagraphical records

- Anni 1992-2001: ISTAT population is reconstructed using census data 1991, 2001 and anagraphical records

- from 2002 on: ISTAT, using the administrative survey entitled "Rilevazione della popolazione residente comunale per sesso, anno di nascita e stato civile al 31 dicembre", based on census data 2001

Note that the indicator is referred to the geographical area of residence.

Population aged 65+; 85+

The population aged 65+ and 85+ in year t, are calculated as follows

%Pop$_{65+}$= (Pop$_{65+}$)/(Pop)*100

%Pop$_{85+}$= (Pop$_{85+}$)/(Pop)*100

Stratification variable is gender (males, females)

Available in: 1982-2015 (annual statistics)

At geographical level: province and region

Source:

- period 1982-1991: ISTAT, population is reconstructed using census data 1981, 1991 and anagraphical records

- Years 1992-2001: ISTAT population is reconstructed using census data 1991, 2001 and anagraphical records

- from 2002 on: ISTAT, using the administrative survey entitled "Rilevazione della popolazione residente comunale per sesso, anno di nascita e stato civile al 31 dicembre" ", based on census data 2001

Note that the indicator is referred to the geographical area of residence.

Foreign residents

There are two indicators of the foreign resident population, calculated as follows:

(a) Count at January 1st

Stratification variable is: gender (males, females)

Available in: 1993-2001, 2003-2016 (annual statistics)

At geographical level: province and region

Source:

Up to 2001: ISTAT, "Bilanci demografici della popolazione residente straniera"

From 2003 on: ISTAT, "Rilevazione sulla Popolazione residente comunale straniera per sesso e anno di nascita"

Note that the indicator s not available for the year 2002 (in this year data collection is interrupted due to the transition from the first to the second survey).

The indicator is referred to the geographical area of the event

(b) Percent distribution of the foreign resident population at 1st January by citizenship

$\%Pop_{FResid}= Pop_{FResid}/(Pop_{TOTResid})*100$

Stratification variable is gender (males, females), citizenship

The considered citizenships are:

Period	Citizenships/Country of birth
	European Union (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Irleand, Luxembourg, The Netherlands, Portugal, UK, Spain, Sweden)
	Central-Eastern Europe (Albania, Bielorussia, Bulgaria, Czech Republic, Cyprus, Croatia, Estonia, Lettonia, Lithuania, Moldavia, Romania, Russia, Slovak Republic, Turkey, Ucraina, Hungary, ex-Jugoslavia)
	North Africa (Algeria, Egypt, Lybia, Morocco, Sudan, Tunisia),
	Western, Eastern and centrer-south Africa (Benin, Burkina Faso, Capo Verde, Ivory Coast, Gambia, Ghana, Guinea, Guinea Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone, Togo, Burundi, Comore, Eritrea, Etiopia, Gibuti, Kenia, Madagascar, Malawi, Mauritius, Mozambico, Ruanda, Seichelles, Somalia, Tanzania, Uganda, Zambia, Zimbabwe, Angola, Botswana, Camerun, central African Republic, Ciad, Congo, Democratic Rep. Congo, Gabon, Guinea, Lesotho, Namibia, Sao Tomé and Prince, Rep. of South Africa, Swaziland)
	East Asia (Brunei, Cambodgia, China, North Korea, South Korea, Philippines, Japan, Indonesia, Laos, Malaysia, Mongolia, Myanmar, Singapore, Taiwan, Thailand, Vietnam)
	West and centrer-south Asia (Saudi Arabia, Armenia, Azerbaigian, Bahrein, United Arab Emirates, Georgia, Giordania, Iran, Iraq, apolyds)
from 2004 to 2006:	European Union (Austria, Belgium, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Lettony, Lithunia, Luxembourg, UK, Malta, The Netherlands, Poland, Portugal, Czech Republic, Slovenia, Slovak republic, Spain, Sweden, Hungary)
	Macedonia, Moldavia, Romania, Serbia and Montenegro, Turkey, Ucraina).
from 2007:	European Union (Austria, Belgium, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Lettony, Lithunia, Luxembourg, UK, Malta, The Netherlands, Poland, Portugal, Czech Republic, Slovenia, Slovak republic, Spain, Sweden, Hungary)
	East-centre Europe centro-orientale (Albania, Bielorussia, Bosnia-Erzegovina, Croatia Russian Federation, Macedonia, Moldavia, Serbia and Montenegro, Turkey, Ucraine).
from 2013:	European Union (Austria, Belgium, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Lettony, Lithunia, Luxembourg, UK, Malta, The Netherlands, Poland, Portugal, Czech Republic, Slovenia, Slovak republic, Spain, Sweden, Hungary)
	East-centre Europe centro-orientale (Albania, Bielorussia, Bosnia-

Period	Citizenships/Country of birth
	Erzegovina, Croatia Russian Federation, Macedonia, Moldavia, Serbia and Montenegro, Turkey, Ucraine).
	European Union (Austria, Belgium, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Lettony, Lithunia, Luxembourg, UK, Malta, The Netherlands, Poland, Portugal, Czech Republic, Slovenia, Slovak republic, Spain, Sweden, Hungary)

Table Annex-2. Country of origin of considered populations

Available in: 1993-2001, 2003-2016 (annual statistics)

At geographical level: province and region

Source:

- up to 2001: ISTAT, "Bilanci demografici della popolazione residente straniera".

- from 2003: ISTAT, "Rilevazione sulla Popolazione residente comunale straniera per sesso e anno di nascita"

Note that the indicator s not available for the year 2002 (in this year data collection is interrupted due to the transition from the first to the second survey).

The indicator is referred to the geographical area of the event

Education

% in the total population of individuals with education level t calculated as:

$$\%P_{edut}=Pop_{edut}/Pop_{tot}*100$$

Stratification variable is gender (males, females), education level (none, primary school, secundary school, high school, university: master degree or PhD)

Available in: 1997-2015 (three-monthly up to 2003; continuous starting from 2004)

At geographical level: province and region

Note that for the years 1997 and 1998, data refer to population aged 6+, for the period 1999-2003 data refer to all ages (0+); starting from 2004 data refer to population aged 15+.

The indicator is referred to the geographical residence area

Occupation
Activity rates

There are 2 activity indicators calculated as follows

$$AR1=FL_{15+}/gP_{15+}*100,$$

$$AR2=FL_{15-64}/P_{15-64}*100$$

Where

FL_{15+}, FL_{15-64} is the active population of age 15+ and 15-64 respectively

P_{15+}, P_{15-64} is the total population of age 15+ and 15-64 respectively

Stratification variable is gender (males, females)

Available for three-monthly intervals up to 2003; continuous starting from 2004:
- 15+: 1993-2014
- 15-64: 1999-2015

At geographical level: province

Source: ISTAT "Indagine sulle forze di lavoro".

The indicator is referred to the geographical area of residence

Employement rates

There are 2 employment rates calculated as follows

$\mathbf{ER_1} = P_{empl,15+}/P_{15+}*100$

$\mathbf{ER_2} = P_{empl,15-64}/P_{15-64}*100$

Where

$P_{empl,15+}$, $P_{empl,15-64}$ is the number of employed persons of age 15+ and 15-64 respectively

P_{15+}, P_{15-64} is the total population of age 15+ and 15-64 respectively

Stratification variable is gender (males, females)

Available for three-monthly intervals up to 2003; continuos starting from 2004: 15+: 1993-2014, 15-64: 1999-2015

At geographical level: province

Source: ISTAT, "Indagine sulle forze di lavoro".

The indicator is referred to the geographical area of residence

Unemployement rates

There are 3 unemployment rates calculated as follows

$\mathbf{UR_1} = P_{unempl,15+}/ FL_{15+}*100$

$\mathbf{UR_2} = P_{unempl,15-64}/ FL_{15-64}*100$

$\mathbf{UR_3} = P_{unempl,15-24}/ FL_{15-24}*100$

Where

$P_{unempl,15+}, P_{unempl,15-64}, P_{unempl,15-24}$ is the number of unemployed persons of age 15+, 15-64 and 15-24 respectively

$FL_{15+}, FL_{15-64}, FL_{15-24}$ is the active population of age 15+, 15-64 and 15-24 respectively

Stratification variable is gender (males, females)

Available for three-monthly intervals up to 2003; continuos starting from 2004; 1993-2015 15+; 1990-2015 for 15-64 and 15-24

At geographical level: province (region for 15-64)

Source: ISTAT, "Indagine sulle forze di lavoro".

The indicator is referred to the geographical area of residence

Poverty

Based on a poverty line (International Standard of Poverty Line - ISPL) defining as poor a household of two components with a consumption expenditure level lower or equal to the mean per-capita consumption expenditure. To define the relative poverty line for different household sizes an equivalence scale is used (Carbonaro equivalence scale) to take into account different needs and economies/diseconomies of scale that can be achieved in bigger/smaller households.

To summarise information on different poverty spread, gravity two indices are calculated:

Incidence of poor persons

Is the proportion of poor persons, calculated as the ratio between the number of individuals in poverty and the total population.

$PI = Pop_{poor}/Pop*100$

Where

Pop_{poor} is the number of individuals in poverty

Pop the total population

Available for three-monthly intervals in 2014

At geographical level: region

Source: ISTAT   "Indagine   sulle   spese   delle   famiglie"
The indicator is referred to the geographical area of residence

Incidence of poor families

Is the proportion of poors calculated as the ratio between the number of households in poverty and the total number of households.

$PH = House_{poor}/House*100$

Where

$House_{poor}$ is the number of households in poverty

House the total number of households

Available for three-monthly interval in: 2002-2014

At geographical level: region

Source:

- Up to 2013: ISTAT, "Indagine sui consumi delle famiglie".

- Starting   from   2014:   ISTAT,   "Indagine   sulle   spese   delle   famiglie".
  The indicator is referred to the geographical area of residence

Note that: since 2014, the data source of poverty estimates (the Household Budget Survey) has been completely reviewed. As a consequence, it has been necessary to reconstruct relative poverty time series since 1997. Time comparisons between 2014 estimates and previously disseminated estimates can be made only using reconstructed data.

Total Fertility rate

The total fertility rate (TF_{15-49}) represents the number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with age-specific fertility rates of the specified year.

Available in: 1980-2014 (annual statistics)

At geographical level: province, region

Source:

- 1980-1998: ISTAT, "Archivio sulla fecondità costituito a partire dalla rilevazione delle nascite".

- Starting from 1999: ISTAT, "Nuova rilevazione degli iscritti in anagrafe per nascita".

Note that: data for Roma province in 2000-2002 are incomplete. For Sardinia provinces TF_{15-49} for the year 2006 refers to the population in 01/01/2007.

The indicator is referred to the geographical area of residence

Birth rate

The Birth Rate is calculated as follows:

BR=Birth$_{alive}$/Pop*1.000

Where

Birth$_{alive}$ is the mnumber of new born alive

Pop is the total population

Available in: 1980-2015 (annual statistics)

At geographical level: province, region

Source:

- 1980-1998: ISTAT, "Archivio sulla fecondità costituito a partire dalla rilevazione delle nascite".

- Starting from 1999: ISTAT, "Nuova rilevazione degli iscritti in anagrafe per nascita".

Note that for the year 1998 the indicator is available at regional level only.

The indicator is referred to the geographical area of residence

Mother's mean age at birth

The indicator, **MeanFx**, provides the mean (average) age of mothers at the birth of their first child.

Available in: 1980-2014 (annual stats)

At geographical level: province, region

Source:

- 1980-1998: ISTAT, "Archivio sulla fecondità costituito a partire dalla rilevazione delle nascite".

- Starting from 1999: ISTAT, "Nuova rilevazione degli iscritti in anagrafe per nascita".

Note that: data for Roma province in 2000-2002 are incomplete. For Sardinia provinces TF_{15-49} for the year 2006 refers to the population in 01/01/2007.

The indicator refers to the geographical area of residence

Abortion

Voluntary abortion

There are the following three indicators of voluntary abortion:

Voluntary abortion rate

VARate= VA counts/PopF*1000

Where

VA counts is the number of voluntary abortion

PopF is the total female population

Available in: 1980-2014 (annual statistics)

At geographical level: province, region

Note that data on VARate are incomplete for the following combinations of region/calendar years:

Piemonte (1986-1995, 1999), Lombardia (2014) Friuli-Venezia Giulia (2005, 2006), Liguria (2013), Umbria (2010, 2012), Marche (2014), Lazio (2005, 2006), Abruzzo (2009, 2012), Molise (2005), Campania (2002, 2005-2014), Puglia (2012, 2013), Basilicata (2009, 2014), Calabria (1981, 1985, 2008), Sicilia (2004-2012, 2014), Sardegna (2008-2009, 2013-2014), Campania Region and provinces (2003).

Age adjusted VARate are calculated using 2001 population as standard.

The indicator refers to the geographical area of residence

Voluntary abortion count

Available in:1980-2014 (annual statistics)

At geographical level: province, region

Note that data on Voluntary abortion counts are incomplete for the following combinations of region/calendar years:

Piemonte (1986-1995, 1999), Lombardia (2014) Friuli-Venezia Giulia (2005, 2006), Liguria (2013), Umbria (2010, 2012), Marche (2014), Lazio (2005, 2006), Abruzzo (2009, 2012), Molise (2005), Campania (2002, 2005-2014), Puglia (2012, 2013), Basilicata (2009, 2014), Calabria (1981, 1985, 2008), Sicilia (2004-2012, 2014), Sardegna (2008-2009, 2013-2014), Campania region and provinces (2003).

The indicator refers to the geographical area of the event

Mother's mean age at voluntary abortion

Available in: 1982-2014(annual statistics)

At geographical level: province, region

Source: ISTAT, "Indagine amministrativa ISTAT sulle interruzioni volontarie della gravidanza"

Note that data on Mother's mean age at voluntary abortion are incomplete for the following combinations of region/calendar years:

Piemonte (1986-1995, 1999), Lombardia (2014) Friuli-Venezia Giulia (2005, 2006), Liguria (2013), Umbria (2010, 2012), Marche (2014), Lazio (2005, 2006), Abruzzo (2009, 2012), Molise (2005), Campania (2002, 2005-2014), Puglia (2012-2013), Basilicata (2009, 2014), Calabria (1981, 1985, 2008), Sicilia (2004-2012, 2014), Sardegna (2008-2009, 2013-2014), Campania Region and provinces (2003).

The indicator refers to the geographical area of residence.

Spontaneous abortion

There are the following three indicators of spontaneous abortion:

SACounts= total number of spontaneous abortions

Available in: 1982-2013 (annual statistics)

At geographical level: province, region

Source: ISTAT:"Indagine amministrativa ISTAT sulle dimissioni dagli istituti di cura per aborto spontaneo"

The indicator refers to the geographical area of the event

Spontaneous abortion Rate

SARate=SAcounts/AliveBirths*1000

Where

Alive Births is the total number of children born alive

Available in: 1982-2013 (annual statistics)

At geographical level: province, region

Source: ISTAT: "Indagine amministrativa ISTAT sulle dimissioni dagli istituti di cura per aborto spontaneo"

Note that data on spontaneous abortion are incomplete for the following combinations of region/calendar years:

Piemonte (1986-1993, 1995-1997), Lombardia (2010), Liguria (2011, 2013), Emilia-Romagna (2013), Toscana (1984), Umbria (2012), Lazio (1995), Abruzzo (2012), Molise (2004, 2011, 2013), all regions (1998).

Age adjusted VARate are calculated using 2001 population as standard.

The indicator refers to the geographical area of the event

Mother's mean age at sp.abortion

Available in: 1982-2013 (annual statistics)

At geographical level: province, region

Source: ISTAT: "Indagine amministrativa ISTAT sulle dimissioni dagli istituti di cura per aborto spontaneo"

Note that data on spontaneous abortion are incomplete for the following combinations of region/calendar years:

Piemonte (1986-1993, 1995-1997), Lombardia (2010), Liguria (2011, 2013), Emilia-Romagna (2013), Toscana (1984), Umbria (2012), Lazio (1995), Abruzzo (2012), Molise (2004, 2011, 2013), all regions (1998).

The indicator refers to the geographical area of the event

2. Mortality-based

Life expectancy at age x: E_x

Where x=0 (at birth), 1, 15, 45, 65, 75

Stratification variable is gender (males, females)

Available in: 1989-2014 (annual statistics)

At geographical level: province, region

Source: ISTAT, "Rilevazione dei decessi.".

Note that, starting from 2004, life expectancy at birth (E_0) is calculated for Piemonte and Valle d'Aosta, Abruzzo and Molise combined. From 1989 to 1992 the indicator is calculated using 5-years data (1985-1989, 1986-90, …) and the year of reference is the upper bound of the interval (1989, 1990, …). Starting from 1996 the indicator is calculated on yearly basis. The indicator refers to the geographical area of residence.

Healthy life expectancy: HE_0

Where x=0 (at birth)

Stratification variable is gender (males, females)

Available in: 2000, 2005 (five-year interval statistics); 2009-2013 (annual statistics)

At geographical level: provinces of Bolzano and Trento, region

Source:

- Up to 2005: ISTAT, "Indagine Multiscopo sulle Famiglie "Condizioni di salute e ricorso ai servizi sanitari".

- Starting from 2009: ISTAT, "Indagine annuale Multiscopo sulle Famiglie "Aspetti della vita quotidiana".

The indicator refers to the geographical area of residence.

Infant deaths

There are two indicators calculated as follows:

Counts <1 year, that is the number of deaths within the first year of life

Stratification variable is gender (males, females)

Available in: 1990-2013 (annual statistics)

At geographical level: province, region

The indicator refers to the geographical area of residence.

Infant mortality rate (IMR)= $Deaths_{<1}/AliveBirths*10.000$

Where

$Deaths_{<1}$ is the number of deaths within the first year of life

AliveBirths is the number of children born alive

Stratification variable is gender (males, females)

Available in: 1990-2013 (annual statistics)

At geographical level: province, region

Source: ISTAT, "Indagine amministrativa ISTAT sulle cause di morte".

The indicator refers to the geographical area of residence.

Stillbirths rate is calculated as follows:

SBR$=Deaths_0/(AliveBirths+ Deaths_0)*10.000$

Where

$Deaths_0$ is the number of deaths at birth

AliveBirths is the number of children born alive

Stratification variable is gender (males, females)

Available in: 1990-2013 (annual statistics)

At geographical level: region

Source:

- Up to 1998: ISTAT, "Archivio sulla fecondità costituito a partire dalla rilevazione delle nascite"

- Starting from 1999

    o Information on Alive Births: ISTAT, "Rilevazione degli iscritti in anagrafe per nascita"

    o Information on deaths at birth: ISTAT, "Movimento naturale della popolazione presente"

Note that the indicator refers to the geographical area of residence up to 1998, and to the geographical area of the event from 1999.

Neonatal deaths

Counts: is the number of deaths within the first 28 days of life.

Stratification variable is gender (males, females)

Available in: 1990-2013 (annual statistics)

At geographical level: province, region

Source: ISTAT, "Indagine amministrativa ISTAT sulle cause di morte". The indicator refers to the geographical area of residence.

Rate

Neonatal mortality rate is calculated as follows:

NMR_{28}= Deaths$_{28days}$/AliveBirths*10.000

Where

Deaths$_{28day}$ is the number of deaths within the first 28 days of life

AliveBirths is the number of children born alive

Stratification variable is gender (males, females)

Available in: 1990-2013 (annual statistics)

At geographical level: province, region

Source: ISTAT, "Indagine amministrativa ISTAT sulle cause di morte". The indicator refers to the geographical area where the event occurs.

Total Mortality:

Counts, crude rates (age specific rates X 10,000) and age standardized rates (std rates X 10,000)

Stratification variable is gender (males, females)

For crude rates (age specific) age groups are: 0-14, 15-24, …, 75+

Available in: 1990-2013 (annual statistics)

At geographical level: province, region

Source: ISTAT, "Indagine amministrativa ISTAT sulle cause di morte".

Note that, for age standadized rates, mean Italian population in 2001 is used as standard.

The indicator refers to the geographical area where the event occurs for counts, and to the residential area for rates (crude and standardized).

Cause specific mortality:

Counts, crude rates (age specific rates X 10,000) and age standardized rates (std rates X 10,000)

Stratification variables are: gender (males, females), cause of death

For crude rates (age specific) age groups are: 0-14, 15-24, …, 75+

Available in:1990-2013 (annual statistics)

At geographical level: province, region

Source: ISTAT, "Indagine amministrativa ISTAT sulle cause di morte".

Note that:

- for age standadized rates, the mean Italian population in 2001 is used as standard

- the indicator refers to the geographical area where the event occurs for counts, and to the residential area for rates (crude and standardized).

- the classification of the causes of death follows the IX revision (ICDO-9) for the period 1980-2002, and the X revision (ICDO-10) starting from 2003.

- the cause of death here considered is the initial cause.

Deaths for drugs addiction

Count and Rate (for people aged 25-44)

Stratification variable is gender (males, females)

Available in:1999-2003, 2006-2013

At geographical level: province, region

Source: ISTAT, "Indagine amministrativa ISTAT sulle cause di morte". The indicator refers to the geographical area where the event occurs for counts, and to the residential area for rates (crude and standardized).

Deaths for alcohol addiction/abuse

Count and Rate (for people aged 25-44)

Stratification variable is gender (males, females)

Available in:

1999-2003, 2006-2013

At geographical level: region

Source: ISTAT, "Indagine amministrativa ISTAT sulle cause di morte".Note that:

- the classification of the causes of death follows the IX revision (ICDO-9) for the period 1980-2002, and the X revision (ICDO-10) starting from 2003;

- the cause of death here considered is the initial cause

3. Morbidity, disability and hospital discharges/activity

Flu Vaccination

Rates, age-specific rates (per 100)

Stratification variables are: gender (males, females), cause of death

For crude rates (age specific) age groups are: 6-23 months, 2-4 years, 5-8,9-14, 15-17, 18-24, 45-64, 65+; 65-74; 75+; overall

Available in: 2004-2015 total rates; 2011-2015 rates stratified by age

At geographical level: region

Source: ISTAT, "Indagine Multiscopo sulle Famiglie "Condizioni di salute e ricorso ai servizi sanitari".

The indicator refers to the residential area.

Infectious and chronic disease

Note that:

- In the year 1997, infectious diseases do not include malaria (information not available);

- In the year 1995, for the region Lazio age specific and age standardized rates are not available, only crude total is available;

Infectious diseases

Counts, age specific rates, age standardized rates

Stratification variable is: gender (males, females)

For crude rates (age specific) age groups are: (0, 1-14, 15-24, …, 75+)

Available in: 1992-2009; 1992-2014 for AIDS

At geographical level: region

Source:

- For the period 1992-1996: ISTAT, "Rilevazione delle notifiche di malattie infettive"

- For the period 1997-2004: ISTAT working on data from the Italian Ministry of Health – "Rilevazione delle notifiche di malattie infettive"

- For AIDS only: Istituto Superiore di Sanità, Centro Operativo AIDS.

Note that:

- Data refer to the residential area for counts and to the area where the event occurs for rates (crude and standardized);

- Infectious diseses data in Molise region might be incomplete;

- Infectious diseses data in Abruzzo region, for the year 2009, might be incomplete (due to the earthquake occurred in 2009);

- Starting from 2013 data on AIDS refer to the diagnosis and not to the notification as before

Chronic diseases

Age specific rates

Stratification variables are: gender (males, females), specific cause (Pulmonary tubercolosis; extrapulmonary tubercolosis; Infectious diarrhea - but non salmonella; Hepatitis A/B ; Salmonellosis non-thify; Malaria; Varicella; Scarlatina; Rubella; Measles; Parotitis; Pertussis; AIDS).

For age specific rates, age groups are: (0, 1-14, 15-24, …, 65+, or otherwise specified)

Available in: 1992-2009; 1992-2013 for AIDS (annual statistics)

At geographical level: region

Source:

- For the period 1992-1996: ISTAT, "Rilevazione delle notifiche di malattie infettive"

- For the period 1997-2004: ISTAT working on data from the Italian Ministry of Health – "Rilevazione delle notifiche di malattie infettive"

- For AIDS only: Istituto Superiore di Sanità, Centro Operativo AIDS.

Hypertension (self-reported)

Prevalence rates (per 100)

Reference population: Residents in a municipality of LHU and registered in the local health office database, aged 18-69 years

Stratification variables are: gender (males, females)

For rates (age specific) age groups are: five-year classes; overall

Available in: 2012-2015 (cumulative period) rates stratified by age

At geographical level: region

Source: PASSI (ISS), "Progressi delle Aziende Sanitarie per la Salute in Italia".

The indicator refers to the persons 18-69 years of age, to the interview date, report that a doctor told them to be hypertensive, that is to have high blood pressure.

Hypercholesterolemia (self-reported)

Prevalence rates (per 100)

Reference population: Residents in a municipality of LHU and registered in the local health office

database, aged 18-69 years

Stratification variables are: gender (males, females)

For rates (age specific) age groups are: five-year classes; overall

Available in: 2012-2015 (cumulative period) rates stratified by age

At geographical level: region

Source: PASSI (ISS), "Progressi delle Aziende Sanitarie per la Salute in Italia".

The indicator refers to the persons 18-69 years of age, to the interview date, report that a doctor told them to be hypercholesterolemic, that is to have high cholesterol in the blood.

Diabetes (self-reported)

Prevalence rates (per 100)

Reference population: Residents in a municipality of LHU and registered in the local health office

database, aged 18-69 years

Stratification variables are: gender (males, females)

For rates (age specific) age groups are: five-year classes; overall

Available in: 2012-2015 (cumulative period) rates stratified by age

At geographical level: region

Source: PASSI (ISS), "Progressi delle Aziende Sanitarie per la Salute in Italia".

The indicator refers to the persons 18-69 years of age to which a doctor has diagnosed diabetes

Depression (symptomps)

Prevalence rates (per 100)

Reference population: Residents in a municipality of LHU and registered in the local health office

database, aged 18-69 years

Stratification variables are: gender (males, females)

For rates (age specific) age groups are: five-year classes; overall

Available in: 2012-2015 (cumulative period) rates stratified by age

At geographical level: region

Source: PASSI (ISS), "Progressi delle Aziende Sanitarie per la Salute in Italia".

The indicator refers to the persons aged 18-69 years who report that they have experienced in the last two weeks of depressed mood symptoms and / or anhedonia permanently.

Health status (self-reported)

Prevalence rates (per 100)

Reference population: Residents in a municipality of LHU and registered in the local health office

database, aged 18-69 years

Stratification variables are: gender (males, females)

For rates (age specific) age groups are: five-year classes; overall

Available in: 2012-2015 (cumulative period) rates stratified by age

At geographical level: region

Source: PASSI (ISS), "Progressi delle Aziende Sanitarie per la Salute in Italia".

The indicator refers to the persons aged 18-69 years, who declare that their current health status is "good" or "very good"

Incidence and prevalence of malignant tumours

Incidence and prevalence of malignant tumours are estimated from the Istituto Superiore di Sanità (ISS) using statistical methods implemented in the software MIAMOD (Mortality and Incidence Analysis MODel, available at http://www.eurocare.it/MiamodPiamod/tabid/60/Default.aspx) and data from the Italian National Institute of Statistics (mortality and population data) and from the Italian population-based cancer Registries (AIRTUM, http://www.registri-tumori.it/cms/en/english-home).

Incidence and prevalence estimates refer to the first malignant primary tumours only, multiple tumours are excluded.

Incidence

Incidence represents the new cancer cases diagnosed in a given population for a given period of time (usually a calendar year)

Count, age specific and age standardized rates (per 100,000)

Stratification variables are: gender (males, females), and cancer site (melanoma, stomach, colon-retto, lung, breast, prostate)

Available in: 1980-2015 (annual statistics)

At geographical level: region

Source: Istituto Superiore di Sanità

Note that:

- National estimates are obtained from a separate model, therefore these estimates might differ from those derived summing up regional estimates;

- All tumours combined do not include ICD-9 173 (non melanoma skin cancer)

Prevalence

Prevalence represents the number of individuals diagnosed with cancer in a population and alive a certain date.

Count and proportion, age standardized prevalence proportion

Stratification variables are: gender (males, females), and cancer site (melanoma, stomach, colon-retto, lung, breast, prostate)

Available in: 1980-2015 (annual statistics)

At geographical level: region

Source: Istituto Superiore di Sanità

Note that:
- National estimates are obtained from a separate model, therefore these estimates might differ from those derived summing up regional estimates;

- All tumours combined do not include ICD-9 173 (non melanoma skin cancer

4. Health care resources

General health care practitioners (GHP)

There are 4 indicators on general health care practitioners (GHP)

GHP Counts

Number of general practitioners

GHP Rate

GHP Rate= GHP/Pop*10.000

Where

GHP is the number of general practitioners

Pop is the total population

% of GHP with 1500+ assisted

GHP_{1500+} Rate= GHP_{1500+}/GHP*100

Where

GHP_{1500+} is the number of general practitioners with more then 1500 assisted.

Resident population and patients per GHP

Calculated as:

Population per GHP=total Pop/GHP counts

Assisted per GHP= total Assisted/ GHP counts

Available in: 1995-2012 (annual statistics)

At geographical level: region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

Note that 1500 patients is the maximum number of assisted per general practitioner permitted according to the current legislation.

Medical oncology doctors

Counts and %

Available in: 2011-2014 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from database "OneKey" Italian Ministry of Health

Note that: data might include age unspecified doctors, thus the total might differ from the sum of age specific counts.

Outpatients and residential facilities

Counts

Stratified by type of facility (outpatient, diagnostic, end of life care facilities)

Available in: 2004-2012 (annual statistics)

At geographical level: region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

Hospital, day-hospital and physical therapy beds

Beds in Hospital

For ordinary admissions, acute diseases, long-term stays and rehabilitations (public hospitals and private accredited hospitals)

BS Counts= number of beds

BSRate= Total beds/Pop*10.000,

BSRate$_{\text{private accredited}}$=Beds in private hospital./Total beds*100

BSUse= number of days of beds use/ number of days of beds availability*100

Stratification variables are: type of activity (ordinary admission, acute diseases, log-tern stay, rehabilitation) and type of hospital (public hospitals and private accredited hospitals)

Available in: 1996-2012 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

For geriatry, gynecology and pediatrics

BSRate$_{\text{public co-payment}}$ = Beds in public hospitals with co-payment/Beds in public hospitals*100

BSRate$_{\text{geriatric}}$= Beds in geriatric units/Pop$_{65+}$*10.000,

BSRate$_{\text{gynecology}}$= Beds in ostetrical/gynecological units/Pop$_{15-49}$*10.000,

BSRate$_{\text{pediatrics}}$= Beds in pediatric units/Pop$_{0-14}$*10.000,

Stratification variable is: type of activity (ordinary admission, acute diseases, log-tern stay, rehabilitation), type of hospital (public hospitals and private accredited hospitals).

Available in: 1996-2012 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

Beds in day hospital: counts and rates

Counts

BDHRate= Total beds in day hospital/Pop*10.000,

%BDH$_{\text{public}}$= Total beds in day hospital in public structures/ Total beds in day hospital *100,

Available in: 1996-2012 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

%DH$_{\text{public}}$=Day Hospital Discharges of acute patients in public hospitals/total Day Hospital Discharges of acute patients *100

Available in: 1999-2014 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

Note that: discharges refer to acute diseses only in public hospitals and private accredited hospitals. Acute diseases do not include Spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units.

DT_{DH}= Therapy duration in day hospital, i.e. number of days in day hospital regimen of acute patients/discharges of acute patients
Available in: 1999-2014 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle schede di dimissione ospedaliera".

Note that: the indicator refers to discharges of patients admitted in day hospital regimen, for acute diseases only, for public hospitals and private accredited hospitals. Acute diseases do not include Spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units. Day Hospital stay includes multiple day hospital admissions (multiple days) with a unique final discharge of the patient. Day hospital stay refers to the total number of treatment days.

Hospital characteristics

Emergency Room (ER)

ERRate=beds in ER/Pop*10.000,

%ER=% individuals admitted in ER/ beds in ER *100,

%ERHosp=Number of Hospitals with ER/Total number of hospitals *100

Available in: 1996-2012 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

Note that, for Sardegna provinces, the indicators in 2006 refer to population at January 1st 2007.

Ambulance type A

AARate =Number of Ambulances type A/Pop*100.000

Available in: 1996-2012 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

Note that, Ambulances type A are those used for emergencies, pediatric ambulances and intensive care ambulances are excluded. For Sardegna provinces, the indicators in 2006 refer to population at January 1st 2007.

Physical therapy

Counts beds: Number of beds in physical therapy units and/or rehabilitation institutes

%BSRes$_{physical\ activity}$ =Number of beds in physical therapy units and/or rehabilitation institutes (residential health care)/Total number of beds in residential health care*100

%BSSemiRes$_{physical\ activity}$ =Number of beds in physical therapy units and/or rehabilitation institutes (partially residential health care)/Total number of beds in partially residential health care*100

Counts of admissions: number of admissions in physical therapy units and/or rehabilitation institutes

Stratification variable: type of health care (residential, partially residential, ambulatorial)

Available in: 1997-2012 (annual statistics)

At geographical level: region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

7. <u>Health care utilization and costs</u>

Hospitalizations and discharges

There are five indicators oh hospitalizations, calculated as follows:

Counts= Number of Hospital admissions

Duration (total days) = Total number of days of hospital stay

Mean duration= Total number of days of hospital stay /Number of hospital admissions

AccessHospitalRate = Total number of days of hospital stay / Total number of days of hospital stay available*100

Rate of hospitalization= Total Number of hospital admissions /Pop*1.000

Stratification variables are: type of activity (acute diseases, long-term stay, rehabilitation), type of hospital (private, public)

Available in: 1996-2012 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle attività gestionali ed economiche delle Unità Sanitarie Locali e delle Aziende Ospedaliere".

Discharges for acute patients

Discharge Rate for Acute diseases

$DARate_x$=Number of Hospital discharges for acute diseases of patients aged $x/Pop_x*10.000$

Stratification variables are: age group (0-14, …, 65-74, 75+), gender (males, females) and type of admission (ordinary, day-hospital)

Available in: 1999-2014 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle schede di dimissione ospedaliera".

Note that: the indicator refers to hospital discharges for acute diseases only, in public hospitals and private accredited hospitals. Acute diseases do not include Spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units. Discharges include patients deceased during hospital stay, and do not include transfer within the same hospital. For Sardegna provinces, the indicators in 2006 refer to population at January 1st 2007.

Discharges by residence for acute patients

There are two indicators calculated as follows:

%DA$_{res}$=DA$_{res}$/DAtot*100

%DA$_{hosp}$=DA$_{hosp}$/DAtot*100

Where

DA$_{res}$ = Number of discharges of patients with acute diseases by residence (province or region)

DA$_{hosp}$ = Number of discharges of patients with acute diseases by type of hospitalization (ordinary admission, day hospital)

DAtot= Total number of discharges of patients with acute diseases

Available in: 1999-2014 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle schede di dimissione ospedaliera".

Note that: the indicators refer to hospital discharges for acute diseases only in public hospitals and private accredited hospitals. Acute diseases do not include Spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units.

Length of stay of Acute disease patients outside the region of residence

%HDAYS$_{outres}$ =number of hospital days for acute patients outside the region of residence/total number of hospital days for acute patients resident in the region

Available in: 1999-2014 (annual statistics)

At geographical level: province, region

Source: ISTAT working on data from the Italian Ministry of Health, "Rilevazione delle schede di dimissione ospedaliera".

Note that: the indicator refers to ordinary admissions with length of stay longer than 3 days (day hospital and admission of 3 days or less are excluded). Acute diseases do not include Spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units.

Discharges for diabetes

%Ddiab =Number of Hospital discharges for diabetes (ICD9 250) of patients aged x/Pop_x*100

Stratification variables are: age group (0-14, …, 65-74, 75+), gender (males, females) and type of admission (ordinary, day-hospital)

Available in: 2001-2015 (annual statistics)

At geographical level: province, region

Source: ISS working on data from the Italian Ministry of Health, "Rilevazione delle schede di dimissione ospedaliera".

Note that: the indicator refers to hospital discharges for diabetes only, in public hospitals and private accredited hospitals. Acute diseases do not include spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units. Discharges include patients deceased during hospital stay, and do not include transfer within the same hospital.

Discharges for asthma

%Dasthma =Number of Hospital discharges for asthma (ICD9 493) of patients aged x/Pop_x*100

Stratification variables are: age group (0-14, …, 65-74, 75+), gender (males, females) and type of admission (ordinary, day-hospital)

Available in: 2001-2015 (annual statistics)

At geographical level: province, region

Source: ISS working on data from the Italian Ministry of Health, "Rilevazione delle schede di dimissione ospedaliera".

Note that: the indicator refers to hospital discharges for asthma only, in public hospitals and private accredited hospitals. Acute diseases do not include spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units. Discharges include patients deceased during hospital stay, and do not include transfer within the same hospital.

Discharges for heart failure

%Dhearth =Number of Hospital discharges for heart failure (ICD9 428) of patients aged x/Pop_x*100

Stratification variables are: age group (0-14, …, 65-74, 75+), gender (males, females) and type of admission (ordinary, day-hospital)

Available in: 2001-2015 (annual statistics)

At geographical level: province, region

Source: ISS working on data from the Italian Ministry of Health, "Rilevazione delle schede di dimissione ospedaliera".

Note that: the indicator refers to hospital discharges for heart failure only, in public hospitals and private accredited hospitals. Acute diseases do not include spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units. Discharges include patients deceased during hospital stay, and do not include transfer within the same hospital.

Discharges for COPD

%Dcopd =Number of Hospital discharges for COPD (ICD9 490-492, 494, 496) of patients aged x/Pop_x*100

Stratification variables are: age group (0-14, …, 65-74, 75+), gender (males, females) and type of admission (ordinary, day-hospital)

Available in: 2001-2015 (annual statistics)

At geographical level: province, region

Source: ISS working on data from the Italian Ministry of Health, "Rilevazione delle schede di dimissione ospedaliera".

Note that: the indicator refers to hospital discharges for COPD only, in public hospitals and private accredited hospitals. Acute diseases do not include spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units. Discharges include patients deceased during hospital stay, and do not include transfer within the same hospital.

Discharges for alcohol related conditions

%Dalcohol =Number of Hospital discharges for alcohol related conditions (ICD9 303) of patients aged x/Pop_x*100

Stratification variables are: age group (0-14, …, 65-74, 75+), gender (males, females) and type of admission (ordinary, day-hospital)

Available in: 2001-2015 (annual statistics)

At geographical level: province, region

Source: ISS working on data from the Italian Ministry of Health, "Rilevazione delle schede di dimissione ospedaliera".

Note that: the indicator refers to hospital discharges for alcohol related conditions only, in public hospitals and private accredited hospitals. Acute diseases do not include spinal cord units, functional rehabilitation, neurological rehabilitation, long-term stay and residual mental hospital units. Discharges include patients deceased during hospital stay, and do not include transfer within the same hospital.

Drugs consumption (covered by NHS)

Gross expenditure

Total

Per capita expenditure = Total expenditure /Pop

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the gross expenditure for drugs covered by NHS

Doses consumed

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for drugs covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for diabetes drugs

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for diabetes drugs (ATC A10) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for anti-hypertensive drugs

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for anti-hypertensive drugs (ATC C02, C03, C07, C08, C09) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for antibiotics drugs

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for antibiotics drugs (ATC J01) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for NSAIDs drugs

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for NSAIDs drugs (ATC M01A) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for drugs for treating osteoporosis

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for drugs for treating osteoporosis (ATC M05B) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for pain drugs

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for pain drugs (ATC N02A, N02BE, N03AX) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for antidepressants drugs

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for antidepressants drugs (ATC N06A) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for anti-asthmatics drugs

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for anti-asthmatics drugs (ATC R03) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for anti-HIV anti-virals drugs

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for anti-HIV anti-virals drugs (ATC J05A) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for Anti-cancer drugs

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for anti-cancer drugs (ATC L01) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Doses consumed for vaccines

Total

DDD per 1000 population per day = Total no. DDD consumed in the period / ((no. inhabitants * days in the period) /1000)

Available in: 2008-2015 (annual statistics)

At geographical level: region

Source: ISS working on data from the Italian Ministry of Health, "Flusso della tracciabilità del farmaco".

Note that: the indicator refers to the number of doses for vaccines (ATC J07) covered by NHS. DDD is the daily maintenance dose for adults for the drug's main indication.

Financial resources

Gross Domestic Product GDP (euro)

Available in: 1980-2014 (annual statistics)

At geographical level: region

Source: ISTAT, "Elaborazioni della Contabilità Nazionale".

Public health spending

There are 7 indicators, calculated as follows:

THCS=Total spending for health care services

THCS$_{capita}$ (Per capita spending for health care services)=THCS/Pop

%TotSpendig$_{function}$= Spending for health care services by type of spending/Total spending for health care services*100

Stratification variable is: type spending (direct health care services, social health care services, pharmaceutical, general practicioner, specialized medical doctors, private services, rehabilitation, other services).

Available in: 1990-2014 (annual statistics)

At geographical level: region

Source: ISTAT, "Elaborazioni della Contabilità Nazionale".

Note that:

- The currency is lira (unit: billion liras) up to 1997, and euro (unit: billion euros) starting from 1998.

- Spending for health care services refers to all public or private accredited services operating directly in the health care system and providing health care services to all citizens. For 2013

the estimates are considered provisional.

.

THCS$_{house}$ (Per household spending for health care services)=THCS/number of households

% THCS$_{House}$= THCS$_{House}$/THCS*100

Available in: 1990-2013 (annual statistics)

At geographical level: region

Source: ISTAT, "Elaborazioni della Contabilità Nazionale su dati dell'indagine ISTAT sui consumi delle famiglie".

Note that estimates for 2013 are considered provisional.

.

% **THCS** gdp= THCS / GDP*100

% THCS$_{House/gdp}$= THCS$_{House}$/GDP*100

Total, public and family spending

Available in: 1990-2013 (annual statistics)

At geographical level: region

Source: ISTAT, "Elaborazioni della Contabilità Nazionale su dati dell'indagine ISTAT sui consumi delle famiglie".

8. Lifestyle

Obesity, overweight

%Obe=Number Obese individuals$_{agex}$/Population$_{agex}$*100

%Over =Number Overweight individuals$_{agex}$/Population$_{agex}$*100

Stratification variables are: gender (males, females) and age group (18+, 18-24, 25-44, 45-64, 65+, age adjusted)

Available in: 2000, 2005 for for people aged 18-24;

1994, 2000, 2002-2014 for all ages

At geographical level: region

Source:

- In 1994, 2000 e 2004: ISTAT, "Indagine Multiscopo sulle Famiglie: Condizioni di salute e ricorso ai servizi sanitari". In 2002, 2003 and

- Starting from 2005: ISTAT, "Indagine Multiscopo sulle Famiglie - Aspetti della vita quotidiana" (annual survey) and "Indagine Multiscopo sulle Famiglie - Condizioni di salute e ricorso ai servizi sanitari" (every 4-5 years survey)

Note that: Obesity is defined as Body Mass Index (BMI= weight in kilograms/height in meters^2) >30, height and weight used here are those self-reported by the individuals. Age standardized indicators (18+) are obtained using 2001 Italian population as standard. The indicators refer to the area (region) of residence.

%Excess weight$_{age6-17}$=Number of obese and overweight individuals$_{age6-17}$/ Population$_{age6-17}$*100

Stratification variable is: gender (males, females)

Available in: 2010-2014 (annual statistics) At geographical level: region

Source: ISTAT, "Indagine Multiscopo sulle Famiglie-Aspetti della vita quotidiana".

Note that: height and weight used here are those self-reported by the individuals, the indicators refer to the area (region) of residence.

Smoking

%Smokers= Smokers$_{agex}$/Population$_{agex}$*100

%heavy smokers=Heavy Smokers$_{agex}$/Smokers$_{agex}$/Population$_{agex}$*100

Stratification variables are: gender (males, females) and age group (15+, 15-24, 25-34, 35-44, 45-54, 55-64, 65+)

Available in: 1993-2003, 2005-2014 (annual statistics)

Period 1993-2003, 2005-2014 (annual stats)

At geographical level: region

Source: ISTAT,"Indagine Multiscopo sulle Famiglie-Aspetti della vita quotidiana". Note that: data are missing for 2004, the indicators refer to the area (region) of residence.

Cigarettes

%DayCig= Total Number of cigarettes smoked per day by individuals$_{agex}$/ Smokers$_{agex}$*100

Stratification variables are: gender (males, females) and age group (15+, 15-24, 25-34, 35-44, 45-54, 55-64, 65+)

Available in: 1993-2003, 2005-2014 (annual statistics)

Period 1993-2003, 2005-2014 (annual stats)

At geographical level: region

Source: ISTAT, "Indagine Multiscopo sulle Famiglie-Aspetti della vita quotidiana". Note that: data are missing for 2004, the indicators refer to the area (region) of residence

Eating habits

%SufficientBreakfast=number of individuals eating sufficient breakfast$_{agex}$/Population$_{agex}$*100

%MainMeal=number of individuals eating main meal$_{agex}$/Population$_{agex}$*100

%Beef_week==number of individuals eating beaf sometimes a week$_{agex}$/Population$_{agex}$*100

%Fish_week==number of individuals eating fish sometimes a week$_{agex}$/Population$_{agex}$*100

%CheeseDay>=1==number of individuals eating cheese more than once a day$_{agex}$/Population$_{agex}$*100

%VegDay>=1==number of individuals eating vegetables more than once a day$_{agex}$/Population$_{agex}$*100

Stratification variables are: gender (males, females) and age group (3+, 3-14, 15-34, 35-44, 45-54, 55-64, 65+)

Available in: 1993-2003,2005-2014 (annual statistics)

At geographical level: region

Source: ISTAT, "Indagine Multiscopo sulle Famiglie-Aspetti della vita quotidiana". Note that: data are missing for 2004, the indicators refer to the area (region) of residence

Breakfast is considered "sufficient" if is something more than coffee/thè only (milk and/or something to eat)

Alcohol consumption

Prevalence rates (per 100)

Reference population: Residents in a municipality of LHU and registered in the local health office database, aged 18-69 years

Stratification variables are: gender (males, females)

For rates (age specific) age groups are: three age classes; overall

Available in: 2012-2015 (cumulative period) rates stratified by age

At geographical level: region

Source: PASSI, "Progressi delle Aziende Sanitarie per la Salute in Italia".

The indicator refers to the persons aged 18-69 years, who in the last 30 days, consumed at least one standard unit of alcohol (Alcohol units is equivalent to 12 grams of ethanol, roughly the amount contained in a can of beer (330 ml), a glass of wine (125 ml) or a shot of liquor (40 ml)).

Physical activity and sports

% regular activity =number of individuals with regular physical activity$_{agex}$/Population$_{agex}$*100

% occasional =number of individuals with occasional physical activity$_{agex}$/Population$_{agex}$*100

% sometimes =number of individuals doing sometimes physical activity$_{agex}$/Population$_{agex}$*100

% never =number of individuals never doing physical activity$_{agex}$/Population$_{agex}$*100

Stratification variables are: gender (males, females) and age group (3+, 3-14, 15-34, 35-44, 45-54, 55-64, 65+)

Available in: 1997-2003,2005-2014 (annual statistics)

At geographical level: region

Source: **ISTAT**, "Indagine Multiscopo sulle Famiglie-Aspetti della vita quotidiana". Note that: data are missing for 2004, the indicators refer to the area (region) of residence

6.1 METADATA OF POPULATION DATA PROVIDED BY ISS

AGGREGATED DEMOGRAPHIC AND SOCIOECONOMIC INDICATORS - ITALY - SOURCE ISTAT (ITALIAN INSTITUTE OF STATISTICS). http://dati.istat.it/

Data can be queried on-line and exported.

NOTE: we have selected and considered the population indicators (aggregated) for which recent data are available (at least from year 2008 onward) and which are aggregated at least at regional level or inferior (province), without aggregation of regions/provinces.

Table Annex-3. Data available for high-resolution study

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
Demographic and socioeconomic	0000-0029; 0043-0048	Resident population		$P_{resident}$; $\%P_{resident65+}$, $\%P_{resident85+}$	total population	province	1982-2015	ten-year	0-4;75+	M,F,M+F		mean: data are missing if population coverage<60%. Considering the mean resident population at a certain year; the indicator is refferred to the geographical area of residence		Istat (HFA)
	0030-0032; 0033-0035; 0036-0038	Dependency ratio		DR_1, DR_2, DR_3	15-64 people x100	province	1982-2015	three groups: 0-14; 65+; 0-14 + 65+		M,F,M+F		considering the mean resident population at a certain year; the indicator is refferred to the geographical area of residence		Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group group	first/last group	gender	Additional stratification			
	0040-0042	Ageing rate		AGrate	0-14 people x100	province	1982-2015	one group: 65+		M,F,M+F		considering the mean resident population at a certain year; the indicator is refferred to the geographical area of residence		Istat (HFA)
	0043-0048	Population aged 65+; 85+		$\%Pop_{65+}$, $\%Pop_{85+}$	total population x100	province	1982-2015	two groups: 65+; 85+		M,F,M+F		considering the mean resident population at a certain year; the indicator is refferred to the geographical area of residence		
	0110-0112,0113-0115, 0120,0130, 0140	Foreign residents		Count at January 1st, %PopF Resid	total resident population x100	province	1993-2001,2003-2016			M,F,M+F	citizenship	the indicator is referred to the geographical area of the event, different geographical classification depending on the period		Istat (HFA)
	0350-0374	Education		$\%P_{edut}$	1997,1998 6+people x100; 1999-2003 total population x100; from 2004 on 15+ people x100.	province	1997-2015			M,F,M+F	level of education	The indicator is referred to the geographical area of residence	1997-2003 three-monthly; from 2004 on yearly	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
	0380-0381, 0390-0391, 0400-0401; 0410-0411, 0420-0421, 0430-0431; 0450, 0460, 0470; 0451, 0461, 0471; 0455, 0465, 0475	Occupation	Activity rates	AR1, AR2	15+ people	province	1993-2015			M,F,M+F		the indicator is referred to the geographical area of residence	1997-2003 three-monthly; from 2004 on yearly	Istat (HFA)
			Employement rates	ER$_1$, ER$_2$	15+ people	province	1993-2015			M,F,M+F		the indicator is referred to the geographical area of residence	1997-2003 three-monthly; from 2004 on yearly	
			Unemployment rates	UR$_1$, UR$_2$, UR$_3$	15+ people; 15-64 people; 15-24 people	province, region for 15-64 rate	1993-2015	15+; 15-64; 15-24		M,F,M+F		the indicator is referred to the geographical area of residence	1997-2003 three-monthly; from 2004 on yearly	
	0480, 0481	Poverty	Incidence of poor persons, households	PI, PH	persons x100; families x100	region	from 2014 on persons; from 2002 on families					the indicator is referred to the geographical area of residence	three-monthly	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
						Geographic coverage	period coverage	age group		gender	Additional stratification			
		Type	Indicator specification	Indicator labels				group	first/last group					
	0180	Fertility	Total fertility rate	TF$_{15-49}$	15-49 female population x1000	province	1980-2014	15-49		F		Data in 2000-2002 are incomplete for the provincia of Rome ; for the province of Sardinia the 2006 incidence is calculated conisdering the 2007 population. The indicator is referred to the geographical area of residence	annual stats	Istat (HFA)
	0190	Birth rate		BR	population x1000	province	1980-1997; 1999-2015					Regional for 1998. The indicator is referred to the geographical area of residence	annual stats	Istat (HFA)
	0210	Mother's mean age		MeanFx	female population x1000	province	1980-2014			F		Data in 2000-2002 are incomplete for the provincia of Rome ; for the province of Sardinia the 2006 incidence is calculated conisdering the 2007 population. The indicator is referred to the geographical area of residence	annual stats	Istat (HFA)
	0220,0221	Abortion	voluntary	VARate	female population x1000	regioni	1980-2014	15-49		F		Data are incomplete see HFA Indicators description	annual stats	Istat (HFA)
				Voluntary abortion count		province	1980-2014					The indicator refers to the geographical area of the event	annual stats	Istat (HFA)
	0233			mother's mean age at		regioni	1982-2014					Data are incomplete see HFA Indicators description	annual stats	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
				voluntary abortion										
	0240,0241		spontaneous	SACounts		province	1982-2013					The indicator refers to the geographical area of the event	annual stats	Istat (HFA)
				SARate	live births x1000	regioni	1982-2013	15-49				Data are incomplete see HFA Indicators description	annual stats	Istat (HFA)
	0260			Mother's mean age at sp.abortion		regioni	1982-2013					Data are incomplete see HFA Indicators description	annual stats	Istat (HFA)
Mortality-based (ICD9; from 2003 on ICD10)	6090,6100	Life expectancy at birth		E_x		province	1989-2014	0,1,15,45,65,75		M,F		In 1989-1992 indicator is referred to quinquennia (1985-1989, 1986-1990, …) .The indicator refers to the geographical area of residence	annual stats	Istat (HFA)
	6110,6120	Healthy live expectancy		HE_0		regioni	2000,2005;2009-2013	0		M,F		Different classification of 'healty life' from 1994 on. The indicator refers to the geographical area of residence	for 2000, 2005 five-year; in 2009-2013 annua	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
													l stats	
	0300-0302; 0303-0305	Infant deaths		count (<1 year), IMR	live births x1000	province	1990-2013			M,F,M+F		The indicator refers to the geographical area of residence	annual stats	Istat (HFA)
	0306-0308	Stillbirths		SBR	stillbirths x1000	regioni	1990-2013			M,F,M+F		The indicator refers to the geographical area of residence up to 1998, from 1999 on refers to the geographical area of the event	annual stats	Istat (HFA)
	0313-0315, 0323-0325, 0333-0335, 0343-0345; 0310-0312, 0320-0322, 0330-0332, 0340-0342	Neonatal deaths		counts: total and by age group(< 1day, 1-6 days, 1-29 days, 1 month+), rate, std.rate	live births x10000	province	1990-2013			M,F,M+F		Rates are referred to the geographical are of the event while deaths are referred to the area of residence	annual stats	Istat (HFA)
Mortality	0270-0272, 0273-0274, 0280-0288, 0290-0298	Mortality		count, age-specific rate, std.rate	population x10000	province	1990-2013	ten-year	0-14;75+	M,F,M+F		std: resident pop.at 2001 in Italy . The indicator is referred to the geographcial area of the event	annual stats	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
Cause specific mortality	from 1000 to 1968	Infectious and parasitic diseases (001-139.8; B19,B25-B99); Tuberculosis (010-018; A15-A19)		count, rate, std.rate	population x10000	province	1990-2003,2006-2013	ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		AIDS (279.1; B20-B24)		count (all and 25-44), rate, std.rate	population x10000	province	1990-2003,2006-2013	25-44; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Tumours and malignant tumours (140-239; C00-D48)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		digestive organs and retroperitoneum/ peritoneum (150-159; C15-C26, C45.1, C48)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		stomach (151; C16)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		colon-rectum-rectosigmoid junction-anus and anal canal (153-154; C18-C21)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
						Geographic coverage	period coverage	age group		gender	Additional stratification			
		Type	Indicator specification	Indicator labels				group	first/last group					
		respiratory and intrathoracic organs (160-165; C30-C39, C45.0, C45.2)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		trachea-bronchus and lung (162; C33-C34)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		breast - female (174; C50)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		cervix and corpus uteri-uterus, part unspecified (179-180, 182; C53-C55)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		prostate (185; C61)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		lymphoid, hematopietic and related tissue (200-208; C81-C96)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Endocrine, nutritional and metabolic diseases (240-278; E00-E90)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group group	age group first/last group	gender	Additional stratification			
		Diabetes mellitus (250; E10-E14)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Diseases of the blood and blood-forming organs, and certain disorders involving the immune mechanism (279-289, but non 279.1; D50-D89)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Diseases of the nervous system and of the sense organs (320-389; G00-H95)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Diseases of the circulatory system (390-459; I00-I99)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Ischaemic heart diseases (410-414; I20-I25)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Cerebrovascular diseases (430-438; I60-I69)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group group	age group first/last group	gender	Additional stratification			
		Diseases of the respiratory system (460-519; J00-J99)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Influenza and pneumonia (480-487; J10-J18)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Chronic lower respiratory diseases (490-496; J40-J47, J67)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Diseases of the digestive system (520-579; K00-K93)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Diseases of liver (571; K70, K73-K74)		count (all and 65+), rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Diseases of the genitourinary system (580-629; N00-N99)		count, rate, std.rate	population x10000	province	1990-2003,2006-2013	classi decenni	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Diseases of the skin and subcutaneous tissue (680-709; L00-L99)		count, rate, std.rate	population x10000	province	1990-2003,2006-2013	classi decenni	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)

D7.2 Collected population data
PROPRIETARY OF BD2Decide CONSORTIUM   91   *16-Jan-17*
Version 5

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
		Diseases of the musculoskeletal system and connective tissue (710-739; M00-M99)		count, rate, std.rate	population x10000	province	1990-2003,2006-2013	classi decenni	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (780-799; R00-R99)		count, rate, std.rate	population x10000	province	1990-2003,2006-2013	classi decenni	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		External causes of morbidity and mortality (E800-E999; V01-Y89)		count, rate, std.rate	population x10000	province	1990-2003,2006-2013	classi decenni	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Transport accidents (E800-E848; V01-V99)		count, rate, std.rate	population x10000	province	1990-2003,2006-2013	classi decenni	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Intentional self-harm (E950-E959; X60-X84)		count, rate, std.rate	population x10000	province	1990-2003,2006-2013	classi decenni	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
		Pregnancy, childbirth and the puerperium (630-676; O00-O99)		detahs (15-49), rate, std.rate		province	1990-2003,2006-2013	15-49; 3 classi:15-24,25-34,3		F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
								5-49						
		Mental and behavioural disorders (290-319; F00-F99)		count, rate, std.rate	population x10000	province	1990-2003,2006-2013	65+; ten-year	0-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
	5510-5515; 5516-5519	Drug addiction		count, rate	population x10000	province	1999-2003, 2006-2013	25-44		M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
	5660-5669	Alcool addiction/abuse		discharges, rate	population x10000	province	1999-2003, 2006-2013	25-44		M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
Morbidity, disability and hospital discharges/activity	**Vaccination**	3330-3338	Flu vaccination		rate; age-specific rate	population x100	region	tot 2004-2015; età 2011-2015	8 groups:6-23 months,2-4years, 5-8years, 9-14years,15-17years,18-44ye		M,F,M+F		Period 2004-2015 for overall and 65+ rates, 2011-2015 for others.	annual stats	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group group	age group first/last group	gender	Additional stratification			
								ars,45-64years,65+years; 65-74, 75+						
Infectious diseases	4180-4219	Infectious diseases		count; rate; std.rate	population x100000	region	1992-2009	ten-year	0; 1-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
	4740-4814	Infectious diseases	Pulmonary tubercolosis ; extrapulmonary tubercolosis	count; rate; std.rate	population x100000	region	1992-2009	ten-year	0; 1-14; 75+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
	4260-4454	Infectious diseases	Infectious diarrhea (but non salmonella) ; Viral Hepatitis; Viral Hepatitis A/B ; Salmonellosis non-thify; Malaria	count; rate; std.rate	population x100000	region	1992-2009	ten-year	0-14;65+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
						Geographic coverage	period coverage	age group		gender	Additional stratification			
		Type	Indicator specification	Indicator labels				group	first/last group					
	4460-4495;4500-4535; 45040-4575:4620-4655;4700-4735; 4820-4855	Infectious diseases	Varicella; Scarlatina; Rubella; Measles; Parotitis; Pertussis	count; rate; std.rate	population x100000	region	1992-2009	total; 0-14; 4 groups: 0,1-4,5-9,10-14 (yrs).		M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
	4860-4896	Infectious diseases	AIDS	count; rate; std.rate	population x100000	region	1992--2014	total; ten-year	0-24;45+	M,F,M+F		1980-2002 ICD9; from 2003 on ICD10. The indicator is referred to the geographcial area of residence	annual stats	Istat (HFA)
Chronic diseases		Hypertension (self-reported)		prevalence rate x 100	18-69 years	region	2012-2015	total; five-year classes		M,F,M+F				PASSI (ISS)
		Hypercholesterolemia (self-reported)		prevalence rate x 100	18-69 years	region	2012-2015	total; five-year classes		M,F,M+F				PASSI (ISS)
		Diabetes (self-reported)		prevalence rate x 100	18-69 years	region	2012-2015	total; five-year classes		M,F,M+F				PASSI (ISS)
		Depression (symptomps)		prevalence rate x 100	18-69 years	region	2012-2015	total; five-year classes		M,F,M+F				PASSI (ISS)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
						Geographic coverage	period coverage	age group		gender	Additional stratification			
		Type	Indicator specification	Indicator labels				group	first/last group					
		Health status (self-reported)		prevalence rate x 100	18-69 years	region	2012-2015	total; five-year classes		M,F,M+F				PASSI (ISS)
Incidence and prevalence of malignant tumours	4900-4997	Melanoma; Stomach; Colon-rettto; Lung; Breast; Prostate		Count, age specific and age standardized rates (per 100,000); prevalence count and proportion, age standardized proportion (per 100,000)	population x100000	region	1980-2015	total (0-99)		M,F,M+F			annual stats	Istat (HFA)
Health care resources	**General and oncological health care**	7010, 7011, 7012, 7013, 7014	General health care	Indicators on general health care practitioners (GHP)	GHP Counts, GHP Rate, % of GHP with 1500+ assisted, Resident		region	1995-2012						Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
				population and patients per GHP										
	9240-9259; 9340-9369	Medical oncology doctors		count, %	population x1000	province	2011-2014							Istat (HFA)
	7051, 7061, 7070	Outpatients and residential facilities		count		region	2004-2012					ambulatory, diagnostic, final life care		Istat (HFA)
Hospital, day-hospital and physical therapy beds	7210-7218, 7220-7228, 7230-7232, 7290-7298	Hospital	Beds in hospital	BSCount, BSRate, BSRate private accredited, BSUse		province	1996-2012				type of activity (ordinary admission, acute diseases, log-tern stay, rehabilitation) , type of hospital (public hospitals and private accredited hospitals)	ordinary admissions, acute diseases, long-term stays and rehabilitations (public hospitals and private accredited hospitals)		Istat (HFA)
	7230-7235, 7240-7241, 7242-7243, 7244-7245, 7246-7247		Beds in hospital	BSRate public co-payment, BSRate geriatric, BSRate gynecology, BSRate pediatrics		province	1996-2012	65+ for geriatry, 15-49 for gynecolo			type of activity (ordinary admission, acute diseases, log-tern stay, rehabilitation) , type of	geriatry, gynecology and pediatrics		Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
						Geographic coverage	period coverage	age group		gender	Additional stratification			
		Type	Indicator specification	Indicator labels				group	first/last group					
								gy			hospital (public hospitals and private accredited hospitals)			
	7250, 7251, 7253, 7257	Day hospital	Day hospital beds	count, BDHRate, %BDH public	population, Total beds in day hospital	province	1996-2012							Istat (HFA)
	7254, 7255		Day hospital discharges	%DH$_{pub lic}$		province	1999-2014							
	7256	Therapy duration in day hospital		DT$_{DH}$		province, region for 15-64 rate	1999-2014							
Hospital characteristics	7390-7394	Hospital characteristics	Emergency Room (ER)	ERRate, %ER, %ERHosp		province, region for 15-64 rate	1996-2012							Istat (HFA)
	7395		Ambulance (type A)	AARate		province, region for 15-64 rate	1996-2012							Istat (HFA)
	7400, 7401, 7402, 7403, 7404	Physical therapy	Institute/centre of physical therapy	Counts beds, %BS$_{Res physical activity}$, %BS$_{Sem iResphysical activity}$		region	1997-2012				type of health care (residential, partially residential, ambulatorial)			Istat (HFA)

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
	7405-7407	Patients (physical therapy)		Counts of admissions		region	1998-2012				type of health care (residential, partially residential, ambulatorial)			Istat (HFA)
Health care utilization and costs	**Hospitalizations and discharges**	7260-7268; 7270-7278; 7280-7288; 7300-7308	hospitalizations		Counts, Duration (total days), Mean duration, Access HospitalRate, Rate of hospitalization	admissions per 1000	province, region	1996-2012				type of activity (acute diseases, long-term stay, rehabilitation), type of hospital (private, public)			Istat (HFA)
	7310-7319, 7320-7329, 7330-7339, 7360-7369, 7370-7379, 7380-7389	discharges	acute patients	DARate$_x$	per 10000	province, region	1999-2014	10-year age classes	0-14, 75+	M, F, M+F	type of admission (ordinary, day-hospital)			Istat (HFA)
	7340-7341, 7342-7343, 7346, 7350-7351;7352-7353; 7344-7345, 7354-7355	discharges by residence	acute patients	%Dares, %Dahosp, %HDAYSoutres	per 100	province, region	1999-2014				type of admission (ordinary, day-hospital)			Istat (HFA)
		discharges for diabetes	acute patients	%Ddiab	per 100	province, region	2001-2015	ten-year classes	0-14, 75+	M, F, M+F	type of admission (ordinary, day-hospital)			Ministry of Health

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
						Geographic coverage	period coverage	age group		gender	Additional stratification			
		Type	Indicator specification	Indicator labels				group	first/last group					
		discharges for asthma	acute patients	%Dasthma	per 100	province, region	2001-2015	ten-year classes	0-14, 75+	M, F, M+F	type of admission (ordinary, day-hospital)			Ministry of Health
		discharges for heart failure	acute patients	%Dhearth	per 100	province, region	2001-2015	ten-year classes	0-14, 75+	M, F, M+F	type of admission (ordinary, day-hospital)			Ministry of Health
		discharges by COPD	acute patients	%Dcopd	per 100	province, region	2001-2015	ten-year classes	0-14, 75+	M, F, M+F	type of admission (ordinary, day-hospital)			Ministry of Health
		discharges by alcohol related conditions	acute patients	%Dalcohol	per 100	province, region	2001-2015	ten-year classes	0-14, 75+	M, F, M+F	type of admission (ordinary, day-hospital)			Ministry of Health
Drugs consumption (covered by NHS)			Gross expenditure	total; per capita	total population	region	2008-2015						annual stats	Ministry of Health
			Doses consumed	total; DDD per 1000 population per day	total population	region	2008-2015						annual stats	Ministry of Health
			Doses consumed for diabetes drugs	total; DDD per 1000 population per day	total population	region	2008-2015						annual stats	Ministry of Health

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
			Doses consumed for anti-hypertensive drugs	total; DDD per 1000 population per day	total population	region	2008-2015						annual stats	Ministry of Health
			Doses consumed for antibiotics drugs	total; DDD per 1000 population per day	total population	region	2008-2015						annual stats	Ministry of Health
			Doses consumed for NSAIDs drugs	total; DDD per 1000 population per day	total population	region	2008-2015						annual stats	Ministry of Health
			Doses consumed for drugs for treating osteoporosis	total; DDD per 1000 population per day	total population	region	2008-2015						annual stats	Ministry of Health
			Doses consumed for pain drugs	total; DDD per 1000 population per day	total population	region	2008-2015						annual stats	Ministry of Health

Background	Indicator Code ISTAT	Indicators			Reference Population	classification							Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification				
								group	first/last group						
			Doses consumed for antidepressants drugs	total; DDD per 1000 population per day	total population	region	2008-2015							annual stats	Ministry of Health
			Doses consumed for anti-asthmatics drugs	total; DDD per 1000 population per day	total population	region	2008-2015							annual stats	Ministry of Health
			Doses consumed for Anti-HIV anti-virals drugs	total; DDD per 1000 population per day	total population	region	2008-2015							annual stats	Ministry of Health
			Doses consumed for Anti-cancer drugs	total; DDD per 1000 population per day	total population	region	2008-2015							annual stats	Ministry of Health
			Doses consumed for vaccines	total; DDD per 1000 population per day	total population	region	2008-2015							annual stats	Ministry of Health
Financi	9040	GDP		GDP		region	1980-								Istat

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group group	first/last group	gender	Additional stratification			
al resources	(HFA)						2014							(HFA)
	9000-9009; 9010-9020; 9021-9029	Public Health Spending	total, % and per capita	THCS, THCScapita, %TotSpendigfunction		region	1990-2014				type of spending			Istat (HFA)
	9030-9032		per household	THCShouse, % THCSHouse		region	1990-2013							Istat (HFA)
	9041--9043		per GDP	% THCS gdp, % THCSHouse/gdp		region	1990-2013							Istat (HFA)
Life styles	2000-2005, 2010-2015, 2020-2025; 2030-2035, 2040-2045, 2050-2055	Anthropometric data	obesity, overweight	%Obe, %Over		region	1994,2000,2002-2014, 2000, 2005 for for people aged 18-24	18+, 25-44, 45-64, 65+		M,F,M+F			survey	
	2060, 2070, 2071		excess weight	%Excess weight$_{age6-17}$		region	2010-2014	6-17		M,F,M+F			survey	
	2080-2086, 2090-2096, 2100-2106; 2110, 2120, 2130	Smoking habits	smokers, heavy smokers	%Smokers, %heavy smokers		region	1993-2003,2005-2014	15+; 10-year age class	15-24,65+	M,F,M+F			survey	

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group group	age group first/last group	gender	Additional stratification			
								es						
	2140-2146, 2150-2156, 2160-2166		cigarettes consume	%DayCig		region	1993-2003,2005-2014	15+; 10-year age classes		M,F,M+F			survey	
	2350-2376	Eating habits	sufficient breakfast	%SufficientBreakfast		region	1993-2003,2005-2014	3+; 10-year age classes	3-14, 65+	M,F,M+F			survey	
	2410, 2420, 2430		main meal	%MainMeal		region	1993-2003,2005-2014	3+		M,F,M+F			survey	
	2440-2466		beef consumption	%Beef_week		region	1993-2003,2005-2014	3+; 10-year age classes	3-14, 65+	M,F,M+F			survey	
	2470-2496		fish consumption	%Fish_week		region	1993-2003,2005-2014	3+; 10-year age classes	3-14, 65+	M,F,M+F			survey	
	2500, 2510, 2520		cheese consumption	%CheeseDay>=1		region	1993-2003,2005-2014	3+		M,F,M+F			survey	
	2530-2556		vegetables consumption	%VegDay>=1		region	1993-2003,2005-2014	3+; 10-year	3-14, 65+	M,F,M+F			survey	

Background	Indicator Code ISTAT	Indicators			Reference Population	classification						Note	type of study	source
		Type	Indicator specification	Indicator labels		Geographic coverage	period coverage	age group		gender	Additional stratification			
								group	first/last group					
								age classes						
		alcohol consumption	Prevalence rates (per 100)		18-69 years	region	2012-2015	total; five-year classes		M,F,M+F			survey	PASSI (ISS)
	2560, 2670, 2680	Physical activity and sports		% regular activity		region	1997-2003,2005-2014	3+; 10-year age classes	3-14, 65+	M,F,M+F			survey	
	2590, 2600, 2610			% occasional		region							survey	
	2620, 2630, 2640			% sometimes		region							survey	
	2650, 2660, 2670			% never		region							survey	