

## 6

## APPENDICE METODOLOGICA

### APPENDIX: METHODOLOGY

#### 6.1. LA BASE DATI

La base dati utilizzata per la realizzazione dell'Atlante è la fonte informativa costruita nell'ambito del progetto per lo studio delle "Differenze socioeconomiche nella mortalità", inserito nel Programma statistico nazionale. La popolazione residente in Italia alla data del 15° Censimento della popolazione e delle abitazioni (2011) è stata seguita nel tempo e sono stati individuati nel periodo di osservazione 2012-2014 tutti gli eventi in uscita per decesso o per trasferimento all'estero. Le fonti principali utilizzate per la costruzione della base dati sono:

- 15° Censimento della popolazione e delle abitazioni
- Indagine su decessi e cause di morte
- Iscrizioni e cancellazioni all'anagrafe per trasferimento di residenza.

È stato effettuato un *record linkage* a livello individuale tra queste fonti utilizzando il codice fiscale come chiave di *linkage*.

Lo studio adotta un approccio di tipo longitudinale, cioè considera la popolazione censita come coorte iniziale e segue nel tempo lo stato in vita, registrando l'eventuale uscita dalla coorte per morte o per trasferimento all'estero.

Tutte le informazioni socio-anagrafiche, inclusa quella relativa al titolo di studio, tanto per le persone decedute quanto per la popolazione complessiva, sono tratte dal Censimento.

Da questa base informativa è, quindi, possibile costruire gli anni-persona per età, titolo di studio, genere e area geografica, nonché calcolare le frequenze dei decessi per cause di morte secondo le stesse caratteristiche demografiche e territoriali.

#### 6.2. LA PERFORMANCE DI RECORD LINKAGE

Le performance di *record linkage* tra i dati di Censimento e quelli di mortalità è risultata molto elevata. La percentuale dei casi abbinati è pari al 97,1% dei decessi avvenuti in Italia

#### 6.1. DATA SOURCE

*The database used for this Atlas was developed by Istat as part of the National statistical programme (PSN) project "Socioeconomic differences in mortality". Adopting a longitudinal approach, the resident population in Italy on the date of the 15th Census of Population and Housing (2011), considered as the initial cohort, was observed over the period 2012-2014 in order to identify all individuals who died or emigrated abroad during those years.*

*The main sources used to build the database were:*

- *15<sup>th</sup> Census of Population and Housing;*
- *Causes of Death Register;*
- *persons registered in and cancelled from the Population Register due to change of residence.*

*To achieve this aim, individual records in the three data sources were linked by using the fiscal code (a unique personal identification number issued to all residents in Italy at birth or upon immigration) as linkage key.*

*All sociodemographic information, including education level, was taken from the Census, either for people who were still alive at the end of follow-up or those who died during the observation period.*

*The resulting database makes it possible to calculate person-years by age, education level, sex, and geographical area, and thus death rates by cause of death according to the same demographic and geographical characteristics.*

#### 6.2. THE PERFORMANCE OF RECORD LINKAGE

*Record linkage performance between Census and mortality data was very good. Indeed, the percentage of matched cases was 97.1% of deaths occurring in Italy in the period 2012-2014 among all individuals residing in Italy on the Census date (table 1). Only deaths occurring among Italian residents registered on*

	DEATHS (NUMBER)	MATCHED DEATHS (NUMBER)	UNMATCHED DEATHS (NUMBER)	MATCH (%)
<b>TOTAL</b>	1 758 529	1 706 733	51 796	97.1
<b>SEX</b>				
Male	848 697	824 875	23 822	97.2
Female	909 832	881 858	27 974	96.9
<b>AGE GROUPS (YEARS)</b>				
0-24	6 212	5 871	341	94.5
25-34	7 203	6 832	371	94.8
35-44	19 804	18 948	856	95.7
45-54	53 637	51 680	1 957	96.4
55-64	115 034	111 168	3 866	96.6
65-74	254 239	246 938	7 301	97.1
75-84	553 619	538 299	15 320	97.2
85-94	640 082	621 734	18 348	97.1
95-104	107 442	104 056	3 386	96.8
105+	1 257	1 207	50	96.0

**Tabella 1.** Numero e percentuale dei decessi\* nel triennio 2012-2014 abbinati al Censimento 2011 per genere e classe di età. Tutte le classi di età.

**Table 1.** Number and percentage of deaths\* occurring in the resident population of Italy during the 3-year period 2012-2014, linked with 2011 Census by sex and age group. All age groups.

\* I decessi sono relativi a popolazione residente e presente in Italia. / Deaths refer to population resident and present in Italy.

nel periodo 2012-2014 e relativi a persone residenti nel Paese alla data del Censimento (tabella 1). Sono stati esclusi dal conteggio delle performance i decessi per i quali è stato possibile stabilire con certezza che tali individui non sono presenti nel database Censimento (principalmente perché erroneamente non censiti o perché entrati in Italia in data successiva al Censimento).

L'analisi delle performance secondo alcune caratteristiche demografiche mostra una bassa variabilità e garantisce, quindi, un'elevata confrontabilità delle misure di mortalità illustrate in questo Atlante, per genere, età e area geografica (tabella 2).

### 6.3. LIVELLO DI ISTRUZIONE

Il livello di istruzione è utilizzato come indicatore della condizione socioeconomica, in quanto fortemente correlato con altre misure di posizione sociale, quali la condizione occupazionale. Il titolo di studio è funzione anche delle condizioni di *early life*, cioè della posizione sociale della famiglia di origine, dell'adozione di determinati stili di vita e delle opportunità di accesso alle cure, e ciò lo rende idoneo per le analisi di sopravvivenza.

Si è scelto di classificare il livello di istruzione nelle 3 modalità: alto, medio e basso, tenendo conto, del titolo di studio conseguito e dell'anno di nascita dell'individuo (tabella 3). Tale scelta è stata fatta per tener conto della riforma del sistema scolasti-

	DEATHS (NUMBER)	MATCHED DEATHS (NUMBER)	UNMATCHED DEATHS (NUMBER)	MATCH (%)
<b>GEOGRAPHICAL AREA OF RESIDENCE</b>				
Nord-Ovest	480.659	467.371	13.288	97,20%
Nord-Est	345.945	339.904	6.041	98,30%
Centro	359.232	345.855	13.377	96,30%
Sud	383.350	369.065	14.285	96,30%
Isole	189.197	184.406	4.791	97,50%
<b>REGION/PROVINCE OF RESIDENCE</b>				
Piemonte	146.210	143.022	3.188	97,80%
Valle d'Aosta	3.799	3.714	85	97,80%
Lombardia	268.526	260.182	8.344	96,90%
Liguria	62.124	60.453	1.671	97,30%
PA Trento	12.014	11.227	787	93,40%
Veneto	136.562	134.286	2.276	98,30%
Friuli Venezia Giulia	41.258	40.568	690	98,30%
Emilia-Romagna	142.162	140.041	2.121	98,50%
PA Bolzano	13.949	13.782	167	98,80%
Toscana	123.843	119.939	3.904	96,80%
Umbria	29.511	28.657	854	97,10%
Marche	49.804	48.764	1.040	97,90%
Lazio	156.074	148.495	7.579	95,10%
Abruzzo	41.700	40.189	1.511	96,40%
Molise	10.408	9.895	513	95,10%
Campania	150.711	145.245	5.466	96,40%
Puglia	107.194	103.105	4.089	96,20%
Basilicata	17.582	16.962	620	96,50%
Calabria	55.755	53.669	2.086	96,30%
Sicilia	144.402	141.192	3.210	97,80%
Sardegna	44.795	43.214	1.581	96,50%
Non definita	146	132	14	89,80%
Totale	1.758.529	1.706.733	51.796	97,10%

**Tabella 2.** Numero e percentuale dei decessi del triennio 2012-2014 abbinati al Censimento 2011 per ripartizione geografica e regione/provincia di residenza. Tutte le classi di età.

**Table 2.** Number and percentage of deaths occurring in the resident population of Italy during the 3-year period 2012-2014, linked with 2011 Census by geographical area and region/province of residence. All age groups.

*the Census date were considered in the computation. Deaths for which it was possible to establish beyond a reasonable doubt that those individuals were not present in the Census (either because they were erroneously not registered or because they had returned to Italy by the date of the subsequent Census) were excluded from the computation. An analysis conducted according to main demographic characteristics showed low variability of record linkage performances, thereby guaranteeing high comparability of mortality measures illustrated in this Atlas by sex, age, and geographical area (table 2).*

EDUCATION LEVEL	YEAR OF BIRTH	CERTIFICATE RECEIVED
LOW	Before 1952	No certificate, completed elementary school
	1952 or over	No certificate, completed elementary school, completed middle school
MEDIUM	Before 1952	Completed middle school
	1952 or over	High school diploma
HIGH	Before 1952	High school diploma
	1952 or over	University degree

**Tabella 3.** Criterio adottato per la classificazione del livello di istruzione in tre modalità.  
**Table 3.** Criterion used for classification of the three education levels.

co del 1962 con la quale fu istituita la scuola media unica, che divenne gratuita e obbligatoria. Per i nati prima del 1952, il livello di istruzione basso include al massimo la licenza elementare, il medio si riferisce alla licenza media e il livello alto include quanti hanno conseguito almeno un diploma di scuola superiore. Per i nati dal 1952, i tre livelli di istruzione si riferiscono rispettivamente a licenza media, licenza superiore e laurea. Questa classificazione del livello d'istruzione su una scala ordinale a tre categorie è motivata dalla necessità di utilizzare una definizione operativa che garantisca la potenza statistica sufficiente per stratificare per genere, ripartizione geografica per le 35 cause e per gruppo di cause di mortalità selezionate.

## 6.4. LE CAUSE DI MORTE SELEZIONATE

Nella tabella 4 sono riportate tutte le cause di morte considerate e i corrispondenti codici della decima revisione della classificazione internazionale delle malattie e dei problemi sanitari correlati (ICD-10), proposta dall'Organizzazione mondiale della sanità.

## 6.5. LA SPERANZA DI VITA PER LIVELLO DI ISTRUZIONE

Le speranze di vita per livello di istruzione sono state calcolate utilizzando le tavole di mortalità.

Le probabilità di morte all'età 0 sono quelle delle tavole di mortalità nazionali e regionali del 2012 pubblicate dall'Istat e disponibili all'indirizzo web demo.istat.it, in quanto non è stato possibile calcolare le probabilità di morte a 0 anni con i dati del *record linkage* mortalità-censimento.

A partire da un anno di età le probabilità di morte sono state calcolate utilizzando i dati di mortalità del *record linkage*. In particolare, è stata seguita una metodologia già adottata dall'i-

CAUSE OF DEATH	ICD CODE
<b>All causes</b>	<b>All codes</b>
<b>Diseases of the circulatory system</b>	<b>I00-I99</b>
Ischaemic heart diseases	I20-I25
Cerebrovascular diseases	I60-I69
<b>Malignant neoplasms</b>	<b>C00-C97</b>
Malignant neoplasms of stomach	C16
Malignant neoplasms of colon, rectum and anus	C18-C21
Malignant neoplasms of liver and intrahepatic bile ducts	C22
Malignant neoplasms of pancreas	C25
Malignant neoplasms of trachea, bronchus and lung	C33-C34
Malignant neoplasms of prostate	C61
Malignant neoplasms of breast	C50
Malignant neoplasms of bladder	C67
Malignant neoplasms of uterus	C53-C55
Malignant neoplasms of ovary	C56
Hodgkin's disease and lymphomas	C81-C86
Leukaemia	C91-C95
Malignant neoplasms of brain and central nervous system	C70-C72
Malignant neoplasms of upper respiratory and digestive tract	C00-C15, C32
<b>Diseases of the respiratory system</b>	<b>J00-J99</b>
Chronic lower respiratory diseases	J40-J47
Pneumonia	J12-J18
<b>Diseases of the genitourinary system</b>	<b>N00-N99</b>
Diseases of kidney and urethra	N00-N29
<b>Diseases of the musculoskeletal system and connective tissue</b>	<b>M00-M99</b>
<b>Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism</b>	<b>D50-D89</b>
<b>Diabetes mellitus</b>	<b>E10-E14</b>
<b>Organic dementia</b>	<b>F01, F03</b>
<b>Parkinson's disease</b>	<b>G20</b>
<b>Alzheimer's disease</b>	<b>G30</b>
<b>Fibrosis and cirrhosis of liver and chronic hepatitis</b>	<b>K70, K73-K74</b>
<b>Accidents</b>	<b>V01-X59, Y85-Y86</b>
Transport accidents	V01-V99, Y85
Accidental falls	W00-W19
<b>Suicide and intentional self-harm</b>	<b>X60-X84, Y87.0</b>

**Tabella 4.** Cause di morte e corrispondenti codici della classificazione ICD-10 (edizione 2009).

**Table 4.** Causes of death and corresponding ICD-10 codes (2009 edition).

## 6.3. EDUCATION LEVEL

*Education level is used as an indicator of socioeconomic status. It strongly correlates with other measures of social status, such as occupational status and social class. Moreover, education level is a function of early life conditions, e.g., the social position of one's family as well as the adoption of a certain lifestyle and ability to access healthcare.*

*Education level has been classified into three categories: high, medium, and low education. To define each level, not only the lev-*

stituto canadese di statistica,<sup>30</sup> che prevede, oltre al calcolo dei tassi di mortalità per classi quinquennali di età, anche la modellizzazione dei tassi di mortalità oltre i 90 anni.

Dalla serie dei tassi di mortalità è stato possibile calcolare le probabilità di morte e, successivamente, anche altre funzioni biometriche delle tavole di mortalità. L'informazione sul titolo di studio viene utilizzata per il calcolo dei rischi di morte esclusivamente nella fascia di età 30-89 anni. Non è stata fatta alcuna distinzione per livello di istruzione nelle fasce d'età <30 anni e ≥90 anni e la misura di mortalità utilizzata è quella calcolata nella popolazione generale. Il limite di età inferiore calcolata (30 anni) è motivato dalla necessità di selezionare individui per i quali il massimo titolo di studio conseguito si può considerare stabile; invece, il limite superiore (90 anni) è suggerito dalla minore frequenza dei decessi e dal fatto che il ruolo dell'istruzione come determinante della mortalità è minore nelle età più avanzate della vita, in cui altri fattori sociali e relazionali possono avere un ruolo più marcato.<sup>31-34</sup>

I tassi di mortalità a 90 anni di età e oltre sono stati modellizzati con il metodo di Kannisto<sup>35</sup> nel seguente modo:

$$\mu_x = \frac{\alpha \cdot e^{\beta x}}{1 + \alpha \cdot e^{\beta x}}$$

dove  $\mu_x$  è la forza di mortalità all'età  $x$  e  $\alpha$  e  $\beta$  sono due parametri da stimare, che rappresentano rispettivamente il livello di mortalità all'età 0 e il tasso di incremento della mortalità per età.

## 6.6. TASSI STANDARDIZZATI DI MORTALITÀ

Per ogni raggruppamento di cause, sono stati calcolati tassi standardizzati per età  $\times 10.000$  anni-persona, per ciascun livello di istruzione, stratificati per genere e regione, applicando i pesi ricavati dalla popolazione standard europea del 2013.<sup>36</sup> I tassi standardizzati si riferiscono al range di età compreso tra i 30 e gli 89 anni.

Sia  $T_x$  il tasso di mortalità specifico per età, dato dal rapporto tra i decessi ( $D_x$ ) e gli anni-persona a rischio ( $AP_x$ ) nel periodo 2012-2014 nella coorte di individui residenti in Italia alla data del Censimento 2011:

$$T_x = \frac{D_x}{AP_x}$$

dove

$$AP_x = \sum_{t=2012}^{2014} P_{x,t} - \frac{1}{2} \cdot (D_{x,t} + T_{x,t})$$

con

$D_{x,t}$  i decessi nella classe di età  $x$  avvenuti nell'anno  $t$ ;  
 $P_{x,t}$  la popolazione nella classe di età  $x$  al 1 gennaio dell'anno  $t$ ;  
 $T_{x,t}$  i trasferimenti all'estero nella classe di età  $x$  e nell'anno  $t$ .

*el attained, but also the individual's year of birth was taken into consideration (table 3). This choice was made to account for the education system reform of 1962, which instituted middle school as free and compulsory. Individuals born before 1952 have low education if they completed at the most elementary school, medium education if they completed at the most middle school and high education if they have at least a high school diploma. Individuals born after 1952 have low, medium and high education if they completed middle school, have a high school diploma or a university degree, respectively.*

*This classification of education level as a 3-category ordinal scale stems from the need to use an operative definition able to guarantee sufficient statistical power to stratify by sex, geographical area, and the 35 selected causes and groups of causes of death.*

## 6.4. SELECTED CAUSES OF DEATH

*Table 4 reports all causes of death considered in the Atlas and their corresponding ICD-10 codes. ICD-10 is the 10<sup>th</sup> revision of the International Classification of Diseases and Related Health Problems proposed by the World Health Organization.*

## 6.5. LIFE EXPECTANCY BY EDUCATION LEVEL

*Life expectancy by education level was calculated using mortality tables.*

*Mortality probability at age 0 was that of the 2012 national and regional mortality tables published by Istat at [demo.istat.it](http://demo.istat.it), as it was not possible to calculate mortality probability at age 0 with the data from the record linkage mortality-Census.*

*Starting from age 1, mortality probability was calculated using mortality data from the record linkage. In particular, the method previously adopted by Statistics Canada<sup>30</sup> was used, which foresees, along with the calculation of mortality rates by 5-year age groups, the estimation of mortality rates over age 90.*

*From the series of mortality rates, it was then possible to derive death probabilities and, subsequently, other biometric functions of the mortality tables as well. Information on attained education level was used to calculate mortality risks exclusively for the age range 30-89 years. On the contrary, no distinction was made for education level in the population aged <30 years and ≥90 years: in these age groups, mortality rates were assumed to be equal to those calculated for the general population across all three different education levels. The choice of the age limit of 30 years was motivated by the need to select individuals for whom the highest education level attained could be objectively considered stable, while that of 90 years by the fact that education level as a determinant of mortality has a smaller role among the elderly, for whom other social and interpersonal factors play a greater role.<sup>31-34</sup>*

*Mortality rates at the age ≥90 years were modelled using the Kannisto<sup>35</sup> method as follows:*

Il tasso standardizzato di mortalità per ciascun livello di istruzione è stato calcolato come segue:

$$Tstd = \frac{\sum_{x=1}^{12} T_x * N_x}{\sum_{x=1}^{12} N_x}$$

dove

$N_x$  è la popolazione standard nella classe di età x-esima;

$x=1$  indica la classe quinquennale 30-35 anni;

$x=12$  la classe 85-89.

Gli intervalli di confidenza al 95% per i tassi standardizzati di mortalità sono stati calcolati ipotizzando l'adattamento alla curva gaussiana

$$IC95\%(Tstd) = \exp\{\log(Tstd) \pm 1,96 * ES(\log(Tstd))\}$$

dove

$$ES(\log(Tstd)) = \sqrt{\frac{\sum \frac{T_x * N_x^2}{n_x}}{(\sum N_x)^2}}$$

$T_x$  sono i tassi specifici per età;

$N_x$  è la popolazione standard nella x-esima classe di età;

$n_x$  è la popolazione in studio nella "x-esima" classe di età.<sup>37</sup>

## 6.7. ANNI DI VITA POTENZIALMENTE PERSI

Gli anni di vita potenzialmente persi all'età x sono calcolati nel seguente modo:

$$YLL_x = (82,6 - x) * d_x$$

dove

$d_x$  è la frequenza dei decessi osservati a ciascuna età x;

82,6 è la speranza di vita alla nascita della popolazione italiana (M+F) nel 2014 (<http://demo.istat.it>). Per gli individui deceduti sopra gli 82,6 anni, il valore utilizzato per il calcolo degli anni di vita potenzialmente persi (*years of life lost*, YLL) è uguale a zero.

È necessario segnalare che la lettura dei risultati deve tenere conto della scelta di un valore unico di riferimento. Il valore YLL indica il numero di anni potenzialmente vivibili e persi prima di raggiungere il limite superiore d'età convenzionalmente adottato, in questo caso a 82,6. La scelta di una soglia unica è ampiamente utilizzata anche in letteratura e gli indicatori vengono utilizzati proprio per confronti territoriali e di genere (<https://stats.oecd.org/>).

La soglia per il calcolo degli YLL avrebbe potuto essere diversificata per genere e in questo caso i valori prodotti sarebbero stati più bassi negli uomini (80-x anziché 82,6-x) e più alti nelle donne (85-x anziché 82,6); tuttavia, la riduzione dei differenziali di genere che si sarebbe osservata negli YLL rispetto a quelli calcolati con una stessa soglia, avrebbe penalizzato pro-

$$\mu_x = \frac{\alpha \cdot e^{\beta x}}{1 + \alpha \cdot e^{\beta x}}$$

where  $\mu_x$  is the force of mortality at age x;

$\alpha$  and  $\beta$  are two estimated parameters that represent the level of mortality at age 0 and the increase in mortality rate per one-point increment in age, respectively.

## 6.6. STANDARDIZED MORTALITY RATES

For each group of causes of death, the age-standardized mortality rates (per 10.000 person-years) were calculated by education level, sex, and region of residence, applying as weights the proportion of individuals in each age group in the European Standard population of 2013.<sup>36</sup> The age-standardized mortality rates refer to ages 30-89.

Let  $T_x$  be the rate by age, given the ratio between deaths ( $D_x$ ) and person-years at risk ( $AP_x$ ) in the period 2012-2014 in the cohort of individuals residing in Italy on the date of the 2011 Census:

$$T_x = \frac{D_x}{AP_x}$$

where

$$AP_x = \sum_{t=2012}^{2014} P_{x,t} - \frac{1}{2} \cdot (D_{x,t} + T_{x,t})$$

with  $D_{x,t}$  the deaths in the age group x occurring in the calendar year t;

$P_{x,t}$  the population in the age group x on 1 January of the calendar year t;

$T_{x,t}$  the emigrations abroad in the age group x, x + Yx and in the calendar year t.

The standardized mortality rate for each education level was calculated as follows:

$$Tstd = \frac{\sum_{x=1}^{12} T_x * N_x}{\sum_{x=1}^{12} N_x}$$

where

$N_x$  is the standard population in the  $i^{th}$  age group;

$x=1$  indicates the 5-year age group 30-35 years;

$x=12$  the age group 85-89.

The 95% confidence intervals for standardized mortality rates were calculated hypothesising the adaption of the Gaussian distribution:

$$IC95\%(Tstd) = \exp\{\log(Tstd) \pm 1,96 * ES(\log(Tstd))\}$$

where

$$ES(\log(Tstd)) = \sqrt{\frac{\sum \frac{T_x * N_x^2}{n_x}}{(\sum N_x)^2}}$$

prio le donne, in quanto più longeve di 5 anni e con un profilo di mortalità più favorevole rispetto a quello degli uomini. Definito YLL<sub>Rx</sub> come il rapporto tra gli anni di vita persi all'età x e l'ammontare della popolazione all'età x (n<sub>x</sub>), il tasso standardizzato per età degli anni di vita potenzialmente persi per ciascuna causa di morte, genere e livello di istruzione è stato calcolato con la seguente formula:

$$YLL\_STDR = \frac{\sum_{x=30}^{x_{rif}} YLL\_R_x * N_x}{\sum_{x=30}^{x_{rif}} N_x} * 10000$$

dove

$x_{rif} = 82,6$ ;

$N_x$  è la popolazione standard per età (in questo caso, è stata utilizzata la popolazione del follow-up censimento-mortalità al 1 gennaio 2012).

## 6.8. RAPPORTO TRA TASSI STANDARDIZZATI DI MORTALITÀ-MORTALITY RATE RATIOS (MRR)

Per quantificare le disuguaglianze socioeconomiche nella mortalità in termini relativi, sono stati stimati i rapporti tra tassi standardizzati di mortalità per età (*mortality rate ratios*, MRR), separatamente per maschi e femmine, considerando come categoria di riferimento il livello di istruzione alto.<sup>37</sup>

$$MRR_m = \frac{Tstd_m}{Tstd_h}$$

$$MRR_l = \frac{Tstd_l}{Tstd_h}$$

dove

h corrisponde al livello di istruzione alto (*high*);

m corrisponde al livello di istruzione medio (*medium*);

l corrisponde al livello d'istruzione basso (*low*).

Gli intervalli di confidenza al 95% per gli MRR sono stati calcolati con la seguente formula:<sup>38</sup>

$$IC95\% (MRR) = \exp\{\log(MRR) \pm 1,96 * ES(\log(MRR))\}$$

Le elaborazioni statistiche sono state effettuate attraverso l'utilizzo di macro Excel e SAS System.

$T_x$  = age-specific rates;

$N_x$  = standard population in the *i*<sup>th</sup> age group;

$n_x$  = population in the study in the *i*<sup>th</sup> age group.<sup>37</sup>

## 6.7 YEARS OF LIFE POTENTIALLY LOST

The years of life potentially lost (YLL) at age x were calculated as follows:

$$YLL_x = (82,6 - x) * d_x$$

where

$d_x$  is the observed death counts at each age x;

82.6 is the life expectancy at birth for the Italian population (M+F) in 2014 (<http://demo.istat.it>). For deceased individuals over the age of 82.6 years, the value used to calculate YLL was equal to zero.

It is necessary to point out that the reading of the results must take into account the choice of a single reference value. The YLL value indicates the number of years potentially liveable and lost before reaching the upper age limit conventionally adopted in this case at 82.6. The choice of a single threshold is also widely used in literature and the indicators are used precisely for territorial and gender comparisons (<https://stats.oecd.org/>).

The threshold for calculating YLLs could have been diversified by gender and in this case the values produced would have been lower in men (80-x instead of 82.6-x) and higher in women (85-x instead of 82.6); however, the reduction of gender differentials that would have been observed in the YLLs compared to those calculated with the same threshold, would have penalized women, the most prolonged of 5 years and with a mortality profile more favourable than that of men.

Defined YLL<sub>Rx</sub> as the rate per age, given by the ratio between years of life lost at age x and the total population at age x (n<sub>x</sub>), the age standardized rates for years of life potentially lost for each cause of death were calculated by sex and education level using the following formula:

$$YLL\_STDR = \frac{\sum_{x=30}^{x_{rif}} YLL\_R_x * N_x}{\sum_{x=30}^{x_{rif}} N_x} * 10000$$

where

$x_{rif} = 82,6$ ;

$N_x$  is the standard population by age (the Census population at the beginning of the follow-up).

## 6.9 FRAZIONE DI MORTALITÀ ATTRIBUIBILE

Per stimare l'impatto del livello di istruzione medio e basso sulla mortalità, è stata calcolata la frazione di mortalità attribuibile (*population attributable fraction*, PAF), separatamente per maschi e femmine, con la seguente formula:

$$PAF = \frac{(p_m * (MRR_m - 1)) + (p_l * (MRR_l - 1))}{p_h + (p_m * MRR_m) + (p_l * MRR_l)} * 100$$

dove

$p_m$  e  $MRR_m$  sono, rispettivamente, la prevalenza e il rapporto tra tassi di mortalità standardizzati per età del livello di istruzione medio;

$p_l$  e  $MRR_l$  sono, rispettivamente, la prevalenza e il rapporto tra tassi di mortalità standardizzati per età del livello di istruzione basso;

$p_h$  è la prevalenza del livello di istruzione alto.<sup>39</sup>

## 6.10 MODELLO AUTOREGRESSIVO BAYESIANO

Le mappe provinciali (n. 110) sono state costruite, separatamente per maschi e femmine, utilizzando tecniche statistiche bayesiane, al fine di tenere conto della bassa numerosità degli eventi che si verificano a livello di piccole aree. Infatti, in queste situazioni, la stima delle misure oggetto di rappresentazione, nel nostro caso i rapporti standardizzati di mortalità (*standardised mortality ratio*, SMR), può risultare inaffidabile a causa della variazione casuale nella distribuzione degli eventi quando la frequenza degli eventi è bassa. Per ciascuna causa, nella stima dell'SMR provinciale i decessi attesi ( $E_j$ ) sono stati ottenuti attraverso la somma per età del prodotto tra gli anni-persona provinciali e i tassi italiani.

Tramite l'applicazione di un modello bayesiano, l'SMR delle province è stato "modellizzato" come dipendente da due componenti di variabilità: l'eterogeneità generale degli SMR tra province e l'eterogeneità degli SMR delle province prossime all'area in esame. L'obiettivo di tale approccio è quello di "lisciare" (*smoothing*) il valore dell'SMR provinciale con i valori medi locali, rendendo l'interpretazione della mappa più robusta. Si deve sempre tenere conto che queste stime possono discostarsi dai valori reali osservati.

In particolare, l'analisi bayesiana è stata condotta utilizzando il modello autoregressivo bayesiano con distribuzione a priori "convoluta". Nello specifico, è stato applicato il modello bayesiano *spatial conditional autoregressive model* (CAR) per piccole aree.<sup>40</sup> I modelli sono stati stimati ipotizzando che:

$$O_j \sim \text{Poisson}(\mu_j)$$

dove  $O_j$  sono gli eventi osservati per ogni provincia italiana  $j$ , per  $j=1, \dots, 110$ .

## 6.8 RATIO BETWEEN STANDARDIZED MORTALITY RATES-MORTALITY RATE RATIOS (MMRs)

To quantify the socioeconomic inequalities in mortality in relative terms, the ratios between age-standardized mortality rates were estimated (mortality rate ratios, MMR) by sex, considering high education level as reference category.<sup>37</sup>

$$MRR_m = \frac{Tstd_m}{Tstd_h}$$

$$MRR_l = \frac{Tstd_l}{Tstd_h}$$

where

$h$  corresponds to high education level;

$m$  corresponds to medium education level;

$l$  corresponds to low education level.

The 95% confidence intervals for the MRRs were calculated as follows:

$$IC95\% (MRR) = \exp\{\log(MRR) \pm 1,96 * ES(\log(MRR))\}$$

Statistical analyses were performed using macro Excel and SAS System.

## 6.9 POPULATION ATTRIBUTABLE FRACTION (PAF)

To estimate the impact of low and medium education level on mortality, the population attributable fraction (PAF) was calculated by sex as follows:

$$PAF = \frac{(p_m * (MRR_m - 1)) + (p_l * (MRR_l - 1))}{p_h + (p_m * MRR_m) + (p_l * MRR_l)} * 100$$

where

$p_m$  and  $MRR_m$  are the prevalence and the mortality rate ratio, respectively, of medium education level;

$p_l$  and  $MRR_l$  are the prevalence and the mortality rate ratio, respectively, of low education level;

$p_h$  is the prevalence of high education level.<sup>39</sup>

## 6.10 BAYESIAN AUTOREGRESSIVE MODEL

The provincial maps (No. 110) were developed by using Bayesian statistical techniques to take into account the low number of events occurring in the small areas. In these situations, the standardized mortality ratios (SMR) may not be reliable due to the random variation of the distribution of events. For each cause, in the estimate of the provincial SMR, expected deaths ( $E_j$ ) were ob-

È stato, quindi, adottato il seguente modello statistico per la stima degli SMR lisciati:

$$\log(\mu_j) = \log(E_j) + a + h_j + c_j$$

che espresso in termini di SMR diventa:

$$SMR_j = \exp(a + h_j + c_j)$$

dove  $h_j$  e  $c_j$  rappresentano, rispettivamente, componenti di eterogeneità spaziale e di clustering, distribuite come una variabile normale la cui precisione è ipotizzata a priori ( $\tau_h$ ,  $\tau_c$ ).

$$h_j \sim CAR(\text{matrice\_province\_adiacenti}_{ij}, w_{ij}, \text{num}_i, \tau_h)$$

dove

$\tau_h \sim \text{Gamma}(0.5, 0.0005)$ ;

$w_{ij} = 1$  se le due province  $i, j$  sono adiacenti, altrimenti 0;

$\text{num}_i =$  numero delle province adiacenti per singola provincia  $i$ .

$c_i \sim \text{Normale}(0, \tau_c)$ ;

dove

$\tau_c \sim \text{Gamma}(0.5, 0.0005)$ .

Le analisi sono state condotte mediante una procedura composta da macro implementate in SAS System 9.3 e supportata dal software Winbugs 1.4.

*tained by the sum per age of the product between provincial person-years and Italian rates. By applying a Bayesian model, the SMR of the provinces by sex was modelled as dependent on 2 sets of variables: the general heterogeneity of the SMRs among provinces and the heterogeneity of the SMRs of the provinces near the area under study. The objective of this approach was to smooth the value of the provincial SMR with local average values, thereby making the results more robust. It must always be taken into account that these estimates may deviate from the actual values observed.*

*The Bayesian analysis was performed using the Bayesian autoregressive model with an a-priori convoluted distribution, applied for small areas.<sup>40</sup> The models were estimated hypothesizing that*

$$O_j \sim \text{Poisson}(\mu_j)$$

*where  $O_j$  are the events observed for each Italian province  $j$ , for  $j=1, \dots, 110$ .*

*The following statistical model was therefore adopted to estimate the smoothed SMR:*

$$\log(\mu_j) = \log(E_j) + a + h_j + c_j$$

*which, expressed in terms of SMR, becomes:*

$$SMR_j = \exp(a + h_j + c_j)$$

*where  $h_j$  and  $c_j$  represent, respectively, components of spatial heterogeneity and of clustering, distributed as a variable Normal, whose precision is hypothesized a priori ( $\tau_h$ ,  $\tau_c$ ).*

$$h_j \sim CAR(\text{neighbourhood\_province\_matrix}_{ij}, w_{ij}, \text{num}_i, \tau_h)$$

*where*

$\tau_h \sim \text{Gamma}(0.5, 0.0005)$ ;

$w_{ij} = 1$  if the two provinces  $i, j$  are adjacent, otherwise 0;

$\text{num}_i =$  number of neighbouring provinces for each province  $i$ .

$c_i \sim \text{Normale}(0, \tau_c)$ ;

*where*

$\tau_c \sim \text{Gamma}(0.5, 0.0005)$ .

*Analyses were performed using macro developed in SAS System 9.3 and supported by Winbugs 1.4. software.*