

istat working papers

N.14
2019

Lo studio longitudinale italiano: integrazione delle indagini sulla salute con dati di mortalità e ospedalizzazione. Metodologia e potenzialità di utilizzo

*Gabriella Sebastiani, Paola Di Filippo, Moreno Demaria, Nicola Caranci,
Lidia Di Minco, Cristina Tamburini, Angelo d'Errico, Francesco Grippo,
Giuseppe Costa*

Direttrice Responsabile:

Patrizia Cacioli

Comitato Scientifico**Presidente:**

Gian Carlo Blangiardo

Componenti:

Corrado Bonifazi	Vittoria Buratta	Ray Chambers	Francesco Maria Chelli
Daniela Cocchi	Giovanni Corrao	Sandro Cruciani	Luca De Benedictis
Gustavo De Santis	Luigi Fabbris	Piero Demetrio Falorsi	Patrizia Farina
Jean-Paul Fitoussi	Maurizio Franzini	Saverio Gazzelloni	Giorgia Giovannetti
Maurizio Lenzerini	Vincenzo Lo Moro	Stefano Menghinello	Roberto Monducci
Gian Paolo Oneto	Roberta Pace	Alessandra Petrucci	Monica Pratesi
Michele Raitano	Giovanna Ranalli	Aldo Rosano	Laura Terzera
Li-Chun Zhang			

Comitato di redazione**Coordinatrice:**

Nadia Mignolli

Componenti:

Ciro Baldi	Patrizia Balzano	Federico Benassi	Giancarlo Bruno
Tania Cappadozzi	Anna Maria Cecchini	Annalisa Cicerchia	Patrizia Collesi
Roberto Colotti	Stefano Costa	Valeria De Martino	Roberta De Santis
Alessandro Faramondi	Francesca Ferrante	Maria Teresa Fiocca	Romina Fraboni
Luisa Franconi	Antonella Guarneri	Anita Guelfi	Fabio Lipizzi
Filippo Moauro	Filippo Oropallo	Alessandro Pallara	Laura Peci
Federica Pintaldi	Maria Rosaria Prisco	Francesca Scambia	Mauro Scanu
Isabella Siciliani	Marina Signore	Francesca Tiero	Angelica Tudini
Francesca Vannucchi	Claudio Vicarelli	Anna Villa	

Cura editoriale:

Vittorio Cioncoloni

Istat Working Papers

Lo studio longitudinale italiano: integrazione delle indagini sulla salute con dati di mortalità e ospedalizzazione. Metodologia e potenzialità di utilizzo

N. 14/2019

ISBN 978-88-458-2001-4

© 2019

Istituto nazionale di statistica
Via Cesare Balbo, 16 – Roma



Salvo diversa indicazione, tutti i contenuti pubblicati sono soggetti alla licenza Creative Commons - Attribuzione - versione 3.0.

<https://creativecommons.org/licenses/by/3.0/it/>

È dunque possibile riprodurre, distribuire, trasmettere e adattare liberamente dati e analisi dell'Istituto nazionale di statistica, anche a scopi commerciali, a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat), marchi registrati e altri contenuti di proprietà di terzi appartengono ai rispettivi proprietari e non possono essere riprodotti senza il loro consenso.

Lo studio longitudinale italiano: integrazione delle indagini sulla salute con dati di mortalità e ospedalizzazione. Metodologia e potenzialità di utilizzo ¹

Gabriella Sebastiani[^], Paola Di Filippo[^], Moreno Demaria^{*}, Nicola Caranci ^{^^} **, Lidia Di Minco[°], Cristina Tamburini[°], Angelo d'Errico^{^^}, Francesco Grippo[^], Giuseppe Costa^{^^}

[^]Istituto Nazionale di Statistica – ISTAT, Servizio Sanità, Salute e Assistenza

[°] Ministero della Salute

^{*} ARPA Piemonte, Epidemiologia Ambientale

^{**} Agenzia Sanitaria e Sociale Regionale, Regione Emilia-Romagna

^{^^} Servizio Epidemiologia Piemonte ASL TO3

Sommario

La mancanza in Italia di indagini sulla salute longitudinali che consentano di studiare l'impatto nel tempo dei determinanti socio-economici e dei fattori di rischio sulle condizioni di salute della popolazione dell'intero Paese, è stata la principale motivazione per attivare il processo di integrazione delle informazioni delle indagini con i flussi di ospedalizzazione e mortalità. Il progetto è stato avviato nel 2005 grazie a una convenzione stipulata tra Istat, Ministero della Salute e Regione Valle d'Aosta, cui successivamente è subentrata la Regione Piemonte. Entrambe le Regioni hanno demandato al Servizio di Epidemiologia (ASL5 Piemonte, attualmente ASL TO3) il compito di curare il progetto per loro conto. Sono stati costruiti due database che integrano le informazioni delle indagini Istat sulla salute condotte nel 2000 e nel 2005 con i ricoveri e i decessi fino al 2014. Obiettivo di questo lavoro è illustrare il processo e la metodologia di costruzione dei database, le strategie di ottimizzazione delle *performance* di *record-linkage* e i controlli di qualità effettuati.

Parole chiave: salute, mortalità, ricoveri, *record-linkage*, diseguaglianze sociali, stili di vita.

Abstract

The lack in Italy of longitudinal health surveys that allow to study the impact of socio-economic determinants and of risk factors on the population health conditions over time and all over the country was the main motivation to activate the process of integrating survey information with hospital admissions and mortality flows. The project was launched in 2005 thanks to an agreement between ISTAT, the Ministry of Health and the Regione Valle d'Aosta, from which the Regione Piemonte subsequently took over. Both regions have delegated the task of carrying out the project on their behalf to the Epidemiology Service (ASL5 Piemonte, currently ASL TO3). Currently, two databases have been built to integrate information from health surveys conducted by Istat in 2000 and 2005 and data on admissions and deaths until 2014. The aim of this work is to illustrate the process and methodology of the database construction, the strategies adopted to improve the record-linkage performance and the quality controls performed.

Keywords: health, mortality, hospitalization, *record-linkage*, social inequality, lifestyles.

¹ La stesura del testo è da attribuire a Gabriella Sebastiani (cap. 1, cap.2), Nicola Caranci (cap.3 par. 3.1), Moreno Demaria (cap.3 par. 3.2 par. 3.3), Lidia Di Minco e Cristina Tamburini (cap. 3 par. 3.4 e cap. 4 par. 4.3 redatti congiuntamente), Paola Di Filippo (cap. 4 par. 4.1 par 4.2 par. 4.4), Francesco Grippo (cap. 5 par. 5.3), Angelo D'Errico (cap. 5 par. 5.5). I par. 5.1, 5.2 e 5.4 del cap. 5 sono stati redatti congiuntamente da Gabriella Sebastiani e Paola Di Filippo. Giuseppe Costa Considerazioni finali.

Si ringraziano Laura Iannucci, Marilena Pappagallo, Elena Demuru, Angelo Lorenti, Manila Bonciani, Lucilla Scarnicchia, Antonella Iorio per l'importante contributo fornito nelle diverse fasi del progetto. Si ringrazia inoltre Sara Fiorini della Top Network per il fondamentale supporto dato nella fase di informatizzazione delle procedure. Il coordinamento complessivo del progetto è di Gabriella Sebastiani.

Indice

	Pag.
1. Il progetto “Differenze di mortalità e di ospedalizzazione secondo lo stato di salute, gli stili di vita e il consumo di servizi sanitari nel campione delle indagini ISTAT sulla salute”	6
1.1 Il contesto	6
1.2 Le finalità del progetto	7
2. Le fonti	7
2.1 L’indagine ISTAT “Condizioni di salute e ricorso ai servizi sanitari”	7
2.2 L’indagine ISTAT “Decessi e cause di morte”	8
2.3 Le Schede di Dimissione Ospedaliera (SDO)	9
2.4 Iscrizioni e cancellazioni all’anagrafe per trasferimento di residenza (ISCAN)	9
3. Il record linkage tra l’indagine sulla salute del 2000 e le altre fonti. La prima sperimentazione e le fasi successive	10
3.1 Acquisizione, trattamento e integrazione delle informazioni anagrafiche degli individui intervistati nell’indagine ISTAT sulla Salute del 2000	10
3.1.1 <i>Controllo delle procedure di registrazione dei dati anagrafici. Il campionamento, l’errore accettabile e l’errore finale</i>	10
3.1.2 <i>Il record linkage tra i dati anagrafici degli stati di famiglia e le informazioni rilevate con l’indagine ISTAT sulla Salute 2000</i>	12
3.2 La prima sperimentazione: il record linkage tra Salute 2000 e i dati di mortalità dal 1999 al 2007	13
3.3 Il record linkage tra Salute 2000 e i dati di mortalità dal 2008 al 2014	15
3.4 Il record linkage con i dati di ospedalizzazione del Ministero della Salute dal 2001 al 2014	16
4. La costruzione del sistema informativo a partire dall’indagine sulla salute del 2005: nuove strategie e metodi	17
4.1 Acquisizione delle informazioni anagrafiche degli individui intervistati per l’indagine ISTAT Salute 2005 e procedure di ricostruzione e validazione dei codici fiscali	17
4.1.1 <i>Correzione dei dati identificativi, calcolo del codice fiscale e verifica del codice fiscale fornito</i>	17
4.1.2 <i>I controlli e la scelta del codice fiscale per il linkage con mortalità e ospedalizzazione</i>	19
4.2 Il record linkage tra indagine salute e archivio mortalità: la creazione dei codici fiscali dei decessi e le nuove procedure	21
4.2.1 <i>La creazione e la scelta del CF dei decessi</i>	22
4.2.2 <i>Il record linkage tra l’indagine Salute 2005 e i decessi 2004-2014</i>	23
4.3 Il record linkage con i dati di ospedalizzazione del Ministero della Salute dal 2004 al 2014	24
4.4 L’aggiornamento del Follow-up 2005 con i trasferimenti all’estero: il record linkage con i dati ISCAN	24

	Pag.
5. Il trattamento e il controllo di qualità dei dati.....	25
5.1 Check e trattamento dei dati	25
5.2 Controllo potenziale bias per copertura parziale del campione	26
5.3 Controllo della <i>performance</i> del linkage con i decessi.....	32
5.4 Controllo della <i>performance</i> del linkage con i ricoveri.....	35
5.5 Confronto tra proporzione di deceduti e ricoverati nei Follow-up delle indagini 1999-2000 e 2004-2005.....	38
6. Considerazioni finali.....	39
7. Appendici	41
Appendice A – Comparazione abbinabili, non abbinabili, campione - Follow-up 2000 e Follow-up 2005	42
Appendice B - Rapporto decessi osservati/attesi per anno, sesso e classi di età - Follow-up 2000 e Follow-up 2005.....	47
Appendice C - Rapporto ricoveri osservati/attesi per anno, sesso e classe di età - Follow-up 2000 e Follow-up 2005.....	48
Riferimenti bibliografici.....	53

1. Il progetto “Differenze di mortalità e di ospedalizzazione secondo lo stato di salute, gli stili di vita e il consumo di servizi sanitari nel campione delle indagini ISTAT sulla salute”

1.1 Il contesto

Il progetto con il quale sono integrati i dati delle indagini sulla salute con i flussi di mortalità e ospedalizzazione nasce nell’ambito di un’ampia gamma di attività di ricerca condotte a livello internazionale e in Italia, che sono volte a studiare le disuguaglianze sociali nella salute (Costa *et al.*, 2014). L’attenzione su questi temi si è accresciuta, negli ultimi anni, per effetto della crisi economica che ha accentuato le disuguaglianze sociali osservate all’interno dei paesi e il divario tra i paesi. La comparazione tra gli stati europei mostra un quadro con disuguaglianze nella salute più pronunciate nei paesi dell’Europa orientale, di medio livello nei paesi dell’Europa nord-occidentale e di minore intensità in quelli dell’area mediterranea (Mackenbach, 2005; Marmot, 2005; Mackenbach *et al.*, 2008; Mackenbach *et al.*, 2016).

Per studiare i differenziali di salute in funzione di indicatori di posizione socio-economica e valutarne l’evoluzione nel tempo, è necessario disporre di basi informative sanitarie che contengano al contempo informazioni sulle condizioni di salute e variabili di condizione socio-economica (istruzione, stato occupazionale, reddito, condizione abitativa, ecc.). Solo raramente le fonti informative sanitarie sono corredate con variabili utili a costruire una stima della condizione socio-economica. In alcuni paesi europei i sistemi informativi sanitari sono strutturati sin dall’origine in modo da consentire l’integrazione dei dati. E’ il caso, ad esempio, della Svezia in cui ciascun individuo ha un identificativo personale che viene utilizzato per tutte le registrazioni.

Per questa ragione, le disuguaglianze nella salute sono state fino ad ora studiate ricorrendo all’integrazione di più fonti di dati. In Italia, salvo alcune eccezioni, quasi mai i sistemi informativi che rilevano fenomeni sanitari sono corredate da informazioni sulla posizione sociale o da altri indicatori di vulnerabilità sociale degli individui ed è necessario quindi integrare i dati sanitari con dati provenienti da altre fonti in cui sono presenti informazioni sulla condizione socio-economica.

Da molti anni, a livello locale sono stati messi a punto sistemi di *record-linkage* di dati di fonte amministrativa (mortalità, ospedalizzazione, certificati di assistenza al parto, ecc.) con fonti, quali anagrafi e censimento, dalle quali è possibile trarre variabili di posizione sociale. Il primo studio di questo tipo è stato condotto a Torino avvalendosi dei dati del censimento del 1971 e successivamente altre città della Toscana e dell’Emilia-Romagna, Venezia e Roma hanno attivato sistemi di osservazione longitudinale fino a giungere alla creazione di una rete nazionale di studi locali della quale sono entrate a far parte altre regioni (Caranci N. *et al.*, 2014, 2018; Stringhini S. *et al.*, 2015.).

Per studiare le disuguaglianze nella salute, a livello nazionale sono state utilizzate le indagini campionarie di popolazione grazie alle quali sono disponibili indicatori dello stato di salute (autovalutazione delle condizioni di salute e presenza di patologie croniche), fattori di rischio legati a comportamenti insalubri e, al contempo, sono presenti variabili di posizione sociale. Il limite di queste fonti è che l’osservazione è di tipo “trasversale”, vale a dire si rilevano contemporaneamente le condizioni di salute e la posizione sociale e non si può quindi stabilire la direzione del nesso di causalità. In altri paesi per queste ed altre ragioni sono stati messi a punto sistemi di indagini sulla salute longitudinali che consentano di superare questo tipo di ostacoli.

L’assenza in Italia di indagini di tipo “longitudinale” ha quindi indotto a cercare una soluzione che consentisse di studiare i differenziali di salute a livello nazionale. La scelta è stata quella di integrare, mediante procedure di *record linkage*, i dati delle indagini sulla salute con flussi di dati correnti relativi a mortalità gestiti dall’Istat (Indagine su decessi e cause di morte) e di ospedalizzazione curati dal Ministero della Salute (Schede di dimissione ospedaliera - SDO). La stessa strategia è stata adottata in numerosi studi condotti in Europa, Canada e Stati Uniti. Tra di essi si possono menzionare gli studi effettuati nell’ambito del progetto EURO-GBD-SE consortium; in partico-

lare l'analisi condotta, in 21 paesi europei, sulle diseguaglianze socio-economiche nella mortalità realizzata attraverso l'integrazione dei flussi di mortalità con le indagini nazionali sulla salute (National Health Surveys NHIS), le indagini di popolazione basate su esami clinici (Health Examination Surveys HES) o le indagini sociali europee (Eikemo *et al.*, 2014) e lo studio sulle differenze per titolo di studio nella speranza di vita libera da disabilità (Mäki N. *et al.*, 2011).

La base del progetto per la costruzione di questo sistema "longitudinale" è rappresentata dalle indagini campionarie, rappresentative della popolazione residente in Italia, condotte dall'Istituto Nazionale di Statistica, a partire dal 1999-2000, sulle condizioni di salute e il ricorso ai servizi sanitari. L'integrazione con i dati di mortalità e ospedalizzazione offre l'opportunità di osservare gli esiti di salute degli individui che hanno partecipato all'indagine e studiare come le caratteristiche individuali e familiari rilevate influiscano, in un periodo di tempo apprezzabile, sulla mortalità e sulla propensione all'ospedalizzazione.

Il progetto "Differenze di mortalità e di ospedalizzazione secondo lo stato di salute, gli stili di vita e il consumo di servizi sanitari nel campione dell'indagine ISTAT sulla salute" è stato avviato nel 2005 grazie a una convenzione stipulata tra ISTAT, Ministero della salute e Regione Valle d'Aosta. Successivamente si è data continuità al progetto con altre convenzioni fino a giungere a quella più recente sottoscritta nel 2015 nella quale alla regione Valle d'Aosta è subentrata la Regione Piemonte. La Regione Valle d'Aosta e, successivamente, la regione Piemonte hanno deman- dato al Servizio di Epidemiologia (ASL5 Piemonte, attualmente ASL TO3) il compito di curare il progetto per loro conto. Per risolvere le problematiche connesse alla protezione dei dati personali il progetto è stato inserito, come studio progettuale, nel Programma Statistico Nazionale. Attualmente è presente nel PSN come Statistica da fonte amministrativa (SDA) con codice IST-02566.

1.2 Le finalità del progetto

Il progetto consente di allineare, almeno in parte, l'Italia ai paesi del Nord Europa che dispon- gono di sistemi di indagine longitudinale sulle traiettorie di vita e di salute di campioni di popola- zione generale o di coorti di nascita.

L'impianto di questa osservazione "longitudinale" risolve infatti molti limiti di selezione insiti nel metodo trasversale di questo tipo di indagini.

Gli obiettivi prioritari del progetto sono:

- studiare le diseguaglianze socio-economiche nella mortalità e nell'ospedalizzazione, gene- rale e causa-specifiche, in base allo stato di salute, agli stili di vita e al consumo di servizi sanitari rilevati con le indagini campionarie ISTAT sulla salute;
- studiare come e in quale misura le caratteristiche rilevate con le indagini influiscano in un periodo di tempo apprezzabile, sulla mortalità e sulla ospedalizzazione;
- valutare l'eterogeneità geografica delle differenze sociali nella mortalità e nella ospedali- zazione;
- fornire informazioni utili per la programmazione e la valutazione in campo preventivo e sanitario.

2. Le fonti

2.1 L'indagine ISTAT "Condizioni di salute e ricorso ai servizi sanitari"

L'indagine campionaria "Condizioni di salute e ricorso ai servizi sanitari" condotta dall'ISTAT ha l'obiettivo di analizzare i comportamenti relativi alla salute e all'utilizzo dei servizi sanitari se- condo le caratteristiche demografiche e socio-economiche dei cittadini. Attraverso le informazioni raccolte direttamente presso gli individui è possibile costruire indicatori su condizioni di salute e qualità della vita, presenza di disabilità, stili di vita, prevenzione, ricorso ai servizi sanitari, uso dei farmaci, ecc. Ciò rende possibile individuare segmenti di popolazione a rischio, studiare le disu- guaglianze nella salute e nell'accesso ai servizi; conoscere i profili degli utilizzatori dei servizi sa- nitari e le modalità di fruizione.

Le indagini sulla salute rappresentano quindi un'utile fonte informativa per la programmazione socio-sanitaria sia a livello nazionale che locale. In particolare nelle ultime edizioni, grazie al contributo del Fondo sanitario nazionale messo a disposizione dal Ministero della salute e dalle Regioni, è possibile effettuare anche analisi sulle condizioni di salute e sull'utilizzo di servizi sanitari a livello locale (aggregati di ASL o grandi ASL). Indagini analoghe sono condotte in tutti i paesi dell'Unione Europea, che ne ha promosso l'armonizzazione per consentire la comparabilità delle informazioni utili alla programmazione di politiche sanitarie comunitarie (European Health Interview Survey – EHIS).

Le indagini sono condotte con cadenza pressoché quinquennale. La popolazione di interesse – ossia l'insieme delle unità statistiche oggetto di investigazione – è costituita dalle famiglie residenti in Italia e dagli individui che le compongono, al netto dei membri permanenti delle convivenze. L'unità di rilevazione è la famiglia di fatto, ossia l'insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi.

L'indagine è condotta mediante quattro distinte rilevazioni a cadenza trimestrale, per tener conto dell'effetto stagionale dei fenomeni, di particolare rilievo per le problematiche della salute. Il disegno di campionamento ha una struttura generale che ricalca quella degli schemi campionari della maggior parte delle indagini sulle famiglie, ossia un disegno a più stadi comuni-famiglie, con stratificazione dei comuni.

Per soddisfare i bisogni informativi a livello territoriale e consentire stime regionali e sub-regionali utili alla programmazione sanitaria locale, per le indagini del 1999/2000 e del 2004/2005, grazie al contributo del Ministero della Salute e delle Regioni, la numerosità campionaria è stata ampliata passando da un campione teorico complessivo nell'anno di 24.000 famiglie ad uno di circa 60.000. Il tasso di risposta delle indagini è molto alto e si attesta all'87,1% nel 2000 e all'83,1% nel 2005.

In ciascuna edizione dell'indagine sono state accolte le istanze informative provenienti dal mondo accademico, epidemiologico e clinico e, avvalendosi delle precedenti esperienze, è stato attuato un progressivo affinamento degli strumenti di rilevazione.

2.2 L'indagine ISTAT "Decessi e cause di morte"

L'Indagine sui decessi e sulle cause di morte rappresenta la più importante fonte esaustiva delle informazioni epidemiologiche del Paese.

L'indagine ha una copertura totale e il suo campo di osservazione è costituito dall'insieme di tutti i decessi che si verificano in un anno di calendario sul territorio nazionale italiano (popolazione presente). L'unità di rilevazione, coincidente con l'unità di analisi, è il singolo individuo deceduto per il quale vengono rilevate informazioni demografiche e di carattere sanitario.

La raccolta dei dati avviene tramite le schede di morte nelle quali il medico che certifica il decesso deve indicare la sequenza morbosa che ha condotto alla morte e gli eventuali altri stati morbosi rilevanti. Le informazioni di carattere demografico e sociale (codice fiscale, sesso, atto di morte, comune e provincia ove è avvenuto il decesso, data di morte, data di nascita, località di nascita, età, stato civile, località di residenza, grado di istruzione, professione, condizione professionale, posizione nella professione, ramo di attività economica, cittadinanza) vengono successivamente riportate dall'ufficiale di Stato civile del comune di decesso.

Le cause di morte si basano sulle informazioni mediche contenute nei certificati di morte che vengono successivamente codificate in base alla ICD-10 (*International Statistical Classification of Diseases, Injuries and Causes of Death, X Revision*) dell'Organizzazione Mondiale della Sanità (OMS). Per ciascun decesso viene individuata la "causa iniziale di morte", ossia "la malattia o il traumatismo che ha avviato il concatenamento degli eventi morbosi che ha condotto direttamente alla morte, oppure l'insieme delle circostanze dell'accidente o della violenza che hanno provocato la lesione traumatica mortale" (definizione adottata dall'OMS). E' alla causa iniziale di morte che fanno riferimento le statistiche di mortalità per causa. Le regole fissate dalla ICD-10 per la selezione della causa iniziale garantiscono la possibilità di effettuare confronti temporali e territoriali. Per la corretta applicazione di tali regole l'ISTAT utilizza le "tavole di decisione Acme" (ed. 2009),

sviluppate e aggiornate dall'NCHS (*National Center for Health Statistics, USA*) per la codifica, manuale e automatica della causa iniziale di morte.

A partire dai decessi del 2003 l'Istat produce e diffonde anche i dati sulle "cause multiple di morte" che consentono di fornire un quadro completo del contesto patologico in cui il decesso è avvenuto, in quanto fanno riferimento al complesso delle cause certificate dal medico al momento del decesso. Tale opportunità di analisi è certamente di grande rilievo nel quadro attuale in cui l'età media alla morte è sempre più avanzata e il ruolo prevalente nella mortalità è giocato dalle malattie cronico-degenerative. In tali circostanze infatti il decesso spesso non è imputabile a una singola patologia bensì a una complessa interazione fra più cause e condizioni che considerate singolarmente potrebbero non essere letali.

I dati annuali sono disponibili come valori assoluti, come tassi di mortalità grezzi e come tassi di mortalità standardizzati. L'uso dei tassi permette di confrontare i livelli di mortalità di popolazioni di dimensioni diverse e con diversa composizione per sesso ed età.

I dati sono utilizzati principalmente per studi epidemiologici (trend, fattori di rischio, ecc.), per prendere decisioni riguardanti le politiche pubbliche in materia di salute e per analizzare gli effetti delle politiche.

2.3 Le Schede di Dimissione Ospedaliera (SDO)

Il flusso informativo delle Schede di Dimissione Ospedaliera (flusso SDO) è lo strumento di raccolta delle informazioni relative a tutti gli episodi di ricovero erogati nelle strutture ospedaliere pubbliche e private presenti nel territorio nazionale.

La banca dati SDO è un elemento fondamentale del Programma Nazionale Esiti (PNE) e l'indagine sui dimessi dagli istituti di cura è inserita nel Programma Statistico Nazionale (SAL-00018).

Le informazioni raccolte comprendono caratteristiche anagrafiche del paziente (fra cui età, sesso, residenza, livello di istruzione), caratteristiche del ricovero (istituto e disciplina di dimissione, regime di ricovero, modalità di dimissione, data di prenotazione, classe di priorità del ricovero) e caratteristiche cliniche (diagnosi principale, diagnosi concomitanti, procedure diagnostiche o terapeutiche).

Nata per finalità di carattere prettamente amministrativo del *setting* ospedaliero, questa banca dati, grazie alla ricchezza di informazioni contenute di carattere amministrativo e clinico, è divenuta un irrinunciabile strumento per un'ampia gamma di analisi ed elaborazioni. Il flusso delle SDO viene, infatti, correntemente impiegato per finalità di carattere economico-gestionale, per studi di carattere clinico-epidemiologico e per analisi a supporto dell'attività di programmazione sanitaria, nonché per il monitoraggio dell'erogazione dei Livelli Essenziali di Assistenza, per la valutazione del rischio clinico ospedaliero, per il calcolo di indicatori di appropriatezza e qualità dell'assistenza erogata, nonché per valutazioni di impatto economico e per il riparto del Fondo Sanitario Nazionale.

2.4 Iscrizioni e cancellazioni all'anagrafe per trasferimento di residenza (ISCAN)

La rilevazione "Iscrizioni e cancellazioni all'anagrafe per trasferimento di residenza" (ISCAN), è condotta dall'Istat con cadenza annuale tramite la raccolta dei modelli di rilevazione che vengono trasmessi dalle anagrafi dei comuni all'Istat telematicamente (in passato anche su supporto cartaceo).

Le iscrizioni riguardano le persone trasferitesi nel comune da altri comuni o dall'estero; le cancellazioni riguardano le persone che si sono trasferite in altro comune o all'estero. Il trasferimento da un comune a un altro decorre dal giorno della richiesta di iscrizione nel comune di nuova residenza, ma viene rilevato quando la pratica migratoria, di ritorno dal comune di cancellazione, risulta definita.

I trasferimenti da e per l'estero sono rilevati nel momento in cui, rispettivamente, viene richiesta l'iscrizione o la cancellazione.

3. Il *record linkage* tra l'indagine sulla salute del 2000 e le altre fonti. La prima sperimentazione e le fasi successive

Per avviare il progetto si è reso necessario, innanzitutto, sperimentare l'utilizzo longitudinale dell'indagine campionaria sulla salute attraverso l'estensione a esiti individuali di mortalità e ospedalizzazione utilizzando tecniche di *record linkage* deterministico già impiegate nell'ambito degli studi locali.

Tale sperimentazione è stata avviata integrando l'archivio dell'indagine ISTAT sulle condizioni di salute e il ricorso ai servizi sanitari 1999-2000 (Salute 2000) con le informazioni dell'indagine sui decessi e le cause di morte dell'ISTAT e con quelle dell'archivio delle schede di dimissione ospedaliera di titolarità del Ministero della Salute (SDO).

3.1 Acquisizione, trattamento e integrazione delle informazioni anagrafiche degli individui intervistati nell'indagine ISTAT sulla Salute del 2000

La chiave di *linkage* utilizzata per la costruzione del database è rappresentata dal Codice Fiscale (CF), disponibile negli archivi di mortalità e ospedalizzazione ma non in quello dell'indagine sulla salute del 2000, nel quale non erano presenti neppure alcune delle informazioni necessarie per la ricostruzione del CF (nome, cognome, comune di nascita).

Il primo passo è stato dunque acquisire le informazioni mancanti dall'archivio cartaceo (conservato dall'ISTAT) degli stati di famiglia forniti dai comuni nel momento in cui avevano partecipato all'indagine. A tale scopo, l'ISTAT ha messo a disposizione del Servizio di Epidemiologia (ASL5 Piemonte, attualmente ASL TO3) questo archivio. Il Servizio di Epidemiologia ha provveduto a registrare su supporto informatico i dati degli individui presenti negli stati di famiglia al momento dell'indagine adottando le misure necessarie a garantire il pieno rispetto della normativa sulla tutela dei dati personali. L'archivio elettronico è stato creato mediante una maschera di immissione dati appositamente costruita in Microsoft-Access, con 11 campi nei quali sono state registrate variabili che si riferiscono all'indagine (trimestre di rilevazione, comune di residenza, identificativo dello stato di famiglia, identificativo dei componenti, indirizzo di residenza, intestatario scheda) e variabili identificative degli intervistati (nome, cognome, comune di nascita, data di nascita, sesso).

I dati degli stati di famiglia cartacei sono risultati molto eterogenei tra i vari comuni che hanno partecipato all'indagine quanto a qualità e completezza. Questa criticità ha indotto il Servizio di Epidemiologia a tenere traccia delle anomalie che si riscontravano di volta in volta. Le anomalie riguardavano singoli individui o interi comuni (es.: mancanza del comune di nascita). Complessivamente si è reso necessario completare le informazioni fornite da circa 100 Comuni che si riferivano a oltre 450 famiglie (complessivamente 1215 individui). Le anomalie riscontrate durante la registrazione sono state riportate nel dettaglio in una relazione che il Servizio di Epidemiologia ha trasmesso all'Istat che ha svolto un'ulteriore lavoro di recupero delle informazioni contattando i comuni interessati. Anche se è stato necessario fare sollecitazioni ripetute e se i tempi di risposta da parte dei comuni in alcuni casi sono stati molto lunghi (da poche settimane a più di 4 mesi), le informazioni mancanti sono state recuperate quasi interamente, con l'eccezione di pochi individui (circa 50).

3.1.1 *Controllo delle procedure di registrazione dei dati anagrafici. Il campionamento, l'errore accettabile e l'errore finale*

Dopo l'immissione dei dati, sono stati condotti dalla ASL 5 Piemonte controlli campionari della qualità dell'archivio degli stati di famiglia registrati fino a fine novembre 2006 (circa il 99,5% dei record individuali).

E' stato estratto un campione casuale che consentisse di rilevare, per ogni operatore dedicato all'immissione dati, un errore pari o superiore all'1% dei campi registrati con una potenza dell'80% (errore di seconda specie, $\beta=0,2$) e un livello di confidenza al 95% (errore di prima specie, $\alpha=0,05$). Sono stati estratti in blocco, con campionamento casuale semplice, 3.232 record (individui) da 1.214 famiglie intervistate.

La procedura di controllo è stata effettuata confrontando a video quanto registrato su supporto cartaceo. I campi su cui si è effettuato il controllo dell'immissione dati sono stati i sei più importanti per l'identificazione (cognome, nome, data di nascita, comune di nascita, sesso, intestatario scheda), per un totale di 19.392 informazioni elementari controllate. La misura della qualità complessiva (senza distinguere per operatore) è risultata buona; l'errore di immissione dei dati è, infatti, di poco superiore all'1% (Tavola 1). La proporzione di campi errati è calcolata sia rispetto alle informazioni che occorrono alla costruzione del CF che al complesso di quelli controllati, di cui si riporta anche la numerosità assoluta.

Ulteriori controlli automatici e manuali sui dati hanno consentito di ridurre allo 0,65% l'errore stimato commesso nella fase di registrazione. Una parte rilevante degli errori di inserimento della data di nascita è stata identificata nella fase successiva in cui l'archivio dei dati anagrafici registrati dagli stati di famiglia è stato linkato con i dati derivanti dal questionario ISTAT.

La procedura per il controllo del nome è stata effettuata tramite la verifica dei nomi che non risultavano nell'archivio dell'anagrafe storica dei residenti a Torino (utilizzando l'ultimo aggiornamento del 2005) che può essere considerato un "thesaurus" con un buon livello di completezza dei nomi in uso in Italia, data anche la presenza consistente di immigrati nella città di Torino provenienti prevalentemente dal Sud del paese. L'utilità del confronto con l'anagrafe storica è stata minore nel caso del controllo dei nomi di immigrati da altri stati. Nei casi in cui il nome registrato dagli stati di famiglia cartacei non corrispondeva con un nome presente nell'anagrafe torinese, si è proceduto ad un controllo puntuale e, in presenza di un errore palese e correggibile senza equivoci, il nome è stato cambiato. Il controllo del dato sul cognome è stato ottenuto tramite il confronto all'interno delle famiglie. Cognomi molto simili ad altri presenti nella famiglia, sono stati modificati rendendoli coerenti con quelli degli altri componenti della famiglia nei casi in cui il cognome degli altri componenti fosse lo stesso per tre o più persone. Il sesso è stato corretto in 2.005 casi e imputato, quando mancante da stato di famiglia, per 764 soggetti. Tale operazione è stata condotta facendo riferimento al nome proprio in tutti i casi in cui il sesso è attribuibile in modo non equivocabile. In sintesi sono stati corretti 380 nomi, 239 cognomi, circa 800 date di nascita e 2.789 record per il sesso.

Tavola 1 – Registrazione dati anagrafici intervistati indagine Salute 1999-2000 - Frequenza percentuale di errore per operatore e campo; percentuale di correzione introdotta ed errore finale della procedura di data-entry, numero campi controllati

Campo dell'archivio anagrafico	Op. 1	Op. 2	Op. 3	Op. 4	Op. 5	Op. 6	Op. 7	Op. 8	Totale	% correzione	Errore finale
Nome	0,14	1,46	0,30	1,00	2,76	0,24	3,57	1,13	1,21	0,27	0,94
Cognome	0,68	0,88	0,30	0,33	2,49	0,00	6,63	0,85	1,49	0,17	1,31
Comune di nascita	0,14	1,17	0,30	0,67	2,21	0,98	1,28	0,00	0,77		0,77
Data di nascita	0,14	0,88	0,59	1,00	0,55	0,00	2,04	0,85	0,68	0,58	0,10
Sesso	0,27	1,17	0,30	0,67	2,49	1,22	3,32	3,97	1,55	1,43	0,11
Intestatario Scheda	0,00	0,00	0,00	0,33	0,28	0,00	1,53	0,57	0,31		0,31
Errore totale (%)	0,23	0,93	0,30	0,67	1,80	0,41	3,06	1,23	1,00		0,65
N° campi controllati	4422	2052	2028	1794	2172	2454	2352	2118	19392		
N° soggetti controllati	737	342	338	299	362	409	392	353	3.232		
Soggetti con info CF errate (%)	1,36	4,68	1,78	4,01	10,22	2,44	13,52	7,08	5,23		5,23
Errore campi CF (%)	0,27	1,11	0,36	0,74	2,10	0,49	3,37	1,36	1,14		1,14
N° campi CF controllati	3685	1710	1690	1495	1810	2045	1960	1765	16160		

Il file dei dati anagrafici è stato sottoposto anche ad altri controlli logico formali. In particolare, per la data di nascita sono stati riscontrati 89 valori anomali per i quali è stata applicata una correzione manuale.

Complessivamente l'archivio così ottenuto contiene i dati anagrafici di 139.991 individui, che

sono i componenti di 52.291 famiglie intervistate.

A seguito delle verifiche e della correzione degli errori del data entry, utilizzando le variabili anagrafiche individuali, è stato creato il Codice Fiscale tramite una procedura automatica implementata nel software SAS.

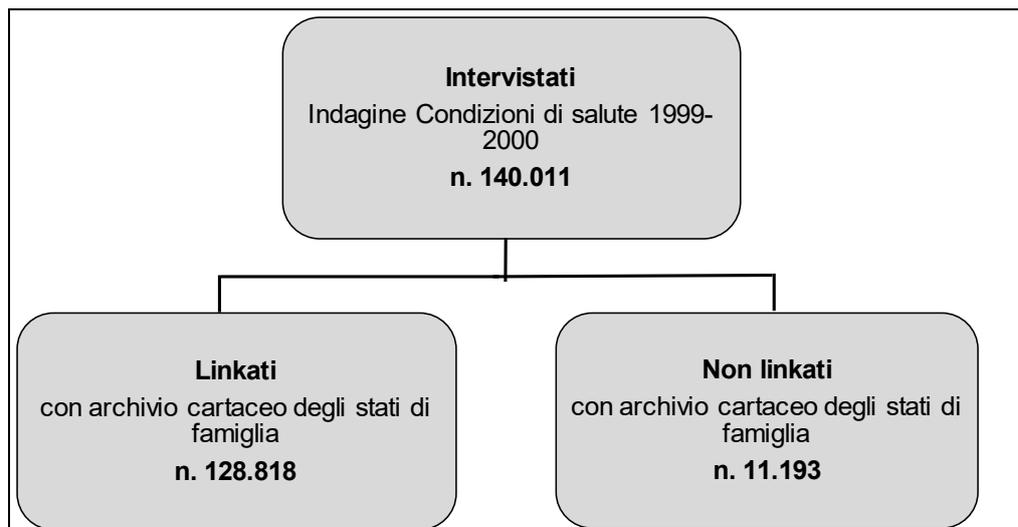
3.1.2 Il record linkage tra i dati anagrafici degli stati di famiglia e le informazioni rilevate con l'indagine ISTAT sulla Salute 2000

Il collegamento dei dati anagrafici degli stati di famiglia di Salute 2000 registrati dall'archivio cartaceo (139.991 individui) con la "coorte" degli intervistati (140.011 individui) è stato effettuato dall'ISTAT mediante una chiave composta dalle seguenti variabili: trimestre di rilevazione, comune di residenza, codice della famiglia, data di nascita e sesso dei componenti.

Nell'effettuare la procedura di *record linkage* deterministico all'archivio dell'indagine salute, è necessario considerare che esiste uno scostamento tra famiglia anagrafica e famiglia di fatto. In altri termini le persone intervistate che non sono presenti negli stati di famiglia non sono incluse nello studio perché non si possono reperire i dati identificativi necessari alla costruzione del CF. Inoltre le persone presenti nello stato di famiglia che non sono state intervistate non fanno parte della coorte degli intervistati oggetto di studio. Tale scostamento è valutato intorno al 5% del campione.

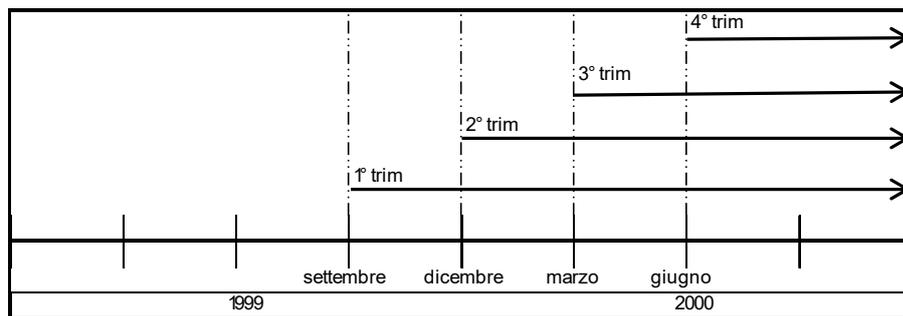
L'unione degli archivi è stata ottenuta con vari passaggi, in cui i record non abbinati sono stati analizzati al fine di correggere, quando possibile, eventuali errori di digitazione nelle variabili della chiave di linkage utilizzata.

Figura 1 – Linkage tra archivio cartaceo stati di famiglia e intervistati Indagine Salute 2000



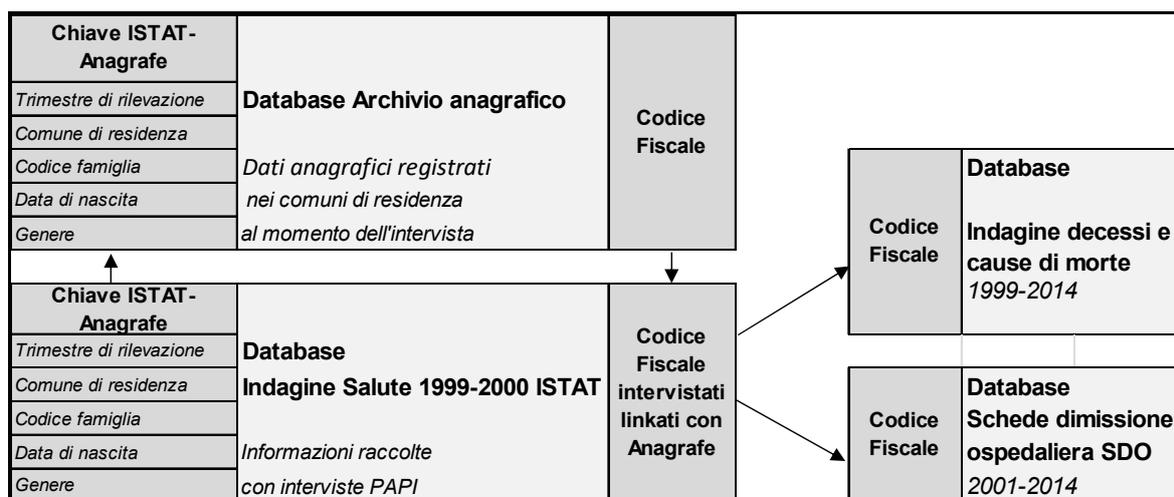
Le procedure di analisi e controllo adottate hanno portato ad un miglioramento della *performance* del *record linkage* deterministico e al raggiungimento di un soddisfacente livello di copertura della "coorte" degli intervistati (140.011 individui). Si è passati infatti da 125.870 record abbinati in una prima fase, a 128.818 record corredati delle informazioni necessarie per il linkage con gli archivi di mortalità e ricoveri (92% degli intervistati nell'indagine Salute 2000). Per una piccola quota di 2.217 intervistati (1,7% sugli abbinabili) non si disponeva delle informazioni relative al cognome (2.215) o al nome (2) e quindi è stato calcolato un CF parziale. L'archivio finale include tutti i 140.011 intervistati, con un campo necessario a distinguere gli 11.193 record non associati tramite la chiave ISTAT.

Figura 2 – Schema concettuale delle fasi della rilevazione dell’indagine Istat Salute 2000 e relativi sviluppi del follow-up nell’osservazione degli esiti sanitari



Per il *record linkage* con i dati di mortalità e ospedalizzazione si è tenuto conto del fatto che la rilevazione per l’indagine Salute 2000 dell’Istat è stata condotta in quattro “trimestri” successivi, a partire dal mese di settembre 1999. Per l’anno 1999 il tempo di osservazione è dunque di 4 quattro mesi per un quarto delle famiglie intervistate e di 1 mese per un altro quarto del campione (famiglie intervistate a dicembre). Per l’anno 2000, per metà delle famiglie (quelle intervistate a settembre e a dicembre 1999) bisogna considerare un anno di osservazione, per un quarto (intervistate a marzo) 10 mesi e per l’ultimo quarto (intervistate a giugno) 7 mesi (Figura 2).

Figura 3 – Schema del data base relazionale generato dall’integrazione delle banche dati che compongono il sistema "campionario longitudinale"



3.2 La prima sperimentazione: il *record linkage* tra Salute 2000 e i dati di mortalità dal 1999 al 2007

L’Istat ha curato l’attività di preparazione degli archivi di mortalità dal 1999 al 2007 e li ha messi a disposizione del Servizio di Epidemiologia della ASL5 Piemonte.

Il CF, che è la chiave di linkage tra i due archivi, è considerato formalmente valido quando è composto da 16 *digit*, contiene solo valori alfanumerici in posizione 1-6, 9, 12 e solo valori numerici nelle restanti posizioni. Inoltre il comune di nascita deve essere valido secondo la tabella dei codici Belfiore/catastali² e in posizione 1-3 e 4-6 non devono essere presenti i caratteri “XXX”, ri-

² Il codice "Belfiore", che corrisponde al codice ISTAT già utilizzato per il codice fiscale personale, è composto da una lettera e tre numeri, e viene assegnato in ordine alfabetico crescente sull'elenco di tutti i comuni di Italia, senza tener conto della Provincia.

velatori della totale assenza di cognome o nome. Questa definizione di validità può essere parziale sia perché in posizioni numeriche possono correttamente trovarsi valori alfanumerici generati da Sogei in caso di omocodia a livello nazionale, sia perché la tabella dei codici catastali dei comuni dovrebbe essere storicizzata. Il CF generato dalla predetta procedura SAS a partire dai dati individuali registrati, ancorché corretti, può quindi differire dal codice rilasciato dal Ministero delle Finanze a causa delle citate omocodie. Nel campione sono attesi 70 casi di possibile omocodia, sicuramente non abbinabili.

Nei file di mortalità 1999-2007 (5.015.166 decessi)³ per ogni record è stato ricalcolato il CF che è risultato concordante con il CF fornito da Istat (dagli archivi di mortalità dell'indagine "Decessi e cause di morte") per il 63,5% dei casi. I CF originali sono risultati presenti e validi nell'81,5% dei casi mentre quelli ricostruiti sono sempre presenti e sono validi per il 95,8% dei casi (Tavola 2).

Tavola 2 – Distribuzione della combinazione degli esiti dei CF forniti e calcolati relativi ai decessi 1999-2007

CF calcolato	CF fornito				Totale	
	Non valido		Valido			
	v.a.	%	v.a.	%	v.a.	%
Non valido	41.942	0,8	167.781	3,3	209.723	4,2
Valido	884.119	17,6	3.921.324	78,2	4.805.443	95,8
TOTALE	926.061	18,5	4.089.105	81,5	5.015.166	100,0

I dati dell'indagine, arricchiti come visto con il codice fiscale ricostruito, sono stati abbinati ai dati di mortalità, contenenti il codice fiscale originale o ricostruito, attraverso una procedura di linkage a passi multipli, ai quali sono stati sottoposti i soli dati dell'indagine non abbinati al passo precedente. Il linkage ha infatti previsto 23 passi (Tavola 3): il primo sul codice fiscale originale o ricostruito, il secondo ha agito utilizzando i soli codici fiscali ricostruiti per rimediare al mancato linkage di CF originali formalmente validi ma errati se confrontati con i dati anagrafici. Dal terzo passo in avanti il linkage ha operato su sotto parti del CF originale o ricostruito di entrambi gli archivi.

Sono utilizzate per primi i CF parzializzati che hanno generato il minor numero di occorrenze multiple negli archivi di mortalità nazionale 1999-2007 comunque inferiori al 2%. Il linkage è sottoposto ad un test di accettazione basato sulla 'distanza' (funzione SAS "Spedis") tra le stringhe formate da cognome e nome presenti nel file di mortalità e nel file dell'indagine sulla Salute 2000.

Le soglie di accettazione sono parzialmente arbitrarie e sono basate sull'analisi a video dei record proposti per l'abbinamento; quest'analisi consente di individuare il valore soglia oltre il quale il record abbinabile risulta sicuramente *mislinked*, ovvero da rifiutare. Dalla Tavola 3 si nota come il maggior contributo al linkage per sotto parti del codice sia dato dalla rimozione del comune (n. key 20), per i noti problemi di attribuzione del codice Belfiore e dal digit 2-3 del nome (n. key 16) poiché in fase di generazione del codice fiscale non è certa la dichiarazione di eventuali secondi nomi.

Le schede di morte degli anni 1999-2007 abbinare ai 128.818 intervistati risultano 8.892, relativamente coerenti secondo i tassi di mortalità attesi nella popolazione generale nel periodo (si veda paragrafo 5.3).

³ I decessi fino al 2003 delle provincie di Trento e Bolzano non sono inclusi in quanto non corredati né delle informazioni relative al nome e cognome, né del CF. I dati non comprendono inoltre i decessi entro il primo anno di vita.

Tavola 3 – Schema e descrizione dei passi della procedura di *record linkage* tra Salute 2000 e decessi 1999-2007

N. key	Linked	% sul totale dei decessi abbinati	Chiave
1	7555	84,96	Con chiave completa originale valida e coerente con dati anagrafici o con chiave ricostruita
2	267	3,00	Con chiave completa ricostruita
3	0	0,00	Senza mese e giorno
4	3	0,03	Senza sesso;
5	32	0,36	Senza digit 2 del nome
6	12	0,13	Senza digit 1 del nome
7	31	0,35	Senza digit 3 del nome
8	17	0,19	Senza mese
9	1	0,01	Senza sesso mese
10	17	0,19	Senza digit 2 del cognome
11	40	0,45	Senza digit 3 del cognome
12	15	0,17	Senza digit 1 del cognome
13	50	0,56	Senza giorno nascita
14	3	0,03	Senza digit 1-2 del nome
15	57	0,64	Senza anno nascita
16	121	1,36	Senza digit 2-3 del nome
17	2	0,02	Senza sesso giorno
18	12	0,13	Senza sesso anno
19	33	0,37	Senza digit 1-3 del nome
20	519	5,84	Senza comune
21	1	0,01	Senza sesso nome
22	6	0,07	Senza sesso comune
23	98	1,10	Senza giorno mese
	8.892	100,0	

3.3 Il *record linkage* tra Salute 2000 e i dati di mortalità dal 2008 al 2014

Il buon esito della procedura sperimentata nella prima fase del progetto, ha consentito di estendere lo studio longitudinale ai decessi dal 2008 al 2014 adottando la medesima strategia affinata e validata grazie all'esperienza acquisita e introducendo alcuni cambiamenti nella preparazione all'abbinamento dei record di mortalità.

Infatti, per gli anni di mortalità dal 2008 al 2014 il codice fiscale è stato sottoposto ad analisi formale e correzione da parte dell'Istat che ha provveduto anche a generare il CF calcolato sulla base dell'algoritmo dell'Agenzia delle Entrate. L'Istat ha inoltre selezionato il CF da utilizzare per il *record linkage* con Salute 2000 sulla base degli esiti delle verifiche effettuate sia sul CF fornito che su quello calcolato. Dal 2013 i dati relativi ai decessi vengono acquisiti già corredati della variabile di scelta del Codice fiscale dall'archivio del progetto ISTAT "Differenze socio-economiche nella mortalità" per il quale sono state applicate le procedure di calcolo e validazione dei CF dei decessi già sperimentate in questo studio⁴.

Al termine delle procedure, dal 2008 al 2014 risultano abbinati 8.915 decessi con un miglioramento dell'abbinamento al primo step (91%) come descritto nella Tavola 4.

⁴ Nell'ambito del progetto "Differenze socio-economiche nella mortalità" si effettua il *record linkage* dei decessi con la popolazione al Censimento 2011. Pertanto le procedure di calcolo e selezione dei CF sono state ottimizzate e standardizzate per consentire l'aggiornamento annuale del database.

Complessivamente i decessi abbinati a Salute 2000 dal 1999 al 2014 sono 17.807 che si riducono a 17.508 in seguito alle procedure di controllo e trattamento effettuate dopo il *record linkage* con le ospedalizzazioni.

Tavola 4 – Schema e descrizione dei passi della procedura di *record linkage* tra Salute 2000 e decessi 2008-2014

N. key	Linked	% sul totale dei decessi	Chiave
1	8113	91,0	Con chiave completa originale valida e coerente con dati anagrafici o con chiave ricostruita
2	0	0,0	Con chiave completa ricostruita
3	0	0,0	Senza mese e giorno
4	4	0,0	Senza sesso;
5	20	0,2	Senza digit 2 del nome
6	6	0,1	Senza digit 1 del nome
7	16	0,2	Senza digit 3 del nome
8	12	0,1	Senza mese
9	2	0,0	Senza sesso mese
10	12	0,1	Senza digit 2 del cognome
11	32	0,4	Senza digit 3 del cognome
12	13	0,1	Senza digit 1 del cognome
13	61	0,7	Senza giorno nascita
14	2	0,0	Senza digit 1-2 del nome
15	52	0,6	Senza anno nascita
16	141	1,6	Senza digit 2-3 del nome
17	5	0,1	Senza sesso giorno
18	24	0,3	Senza sesso anno
19	2	0,0	Senza digit 1-3 del nome
20	234	2,6	Senza comune
21	2	0,0	Senza sesso nome
22	2	0,0	Senza sesso comune
23	160	1,8	Senza giorno mese
	8915	100,0	

3.4 Il *record linkage* con i dati di ospedalizzazione del Ministero della Salute dal 2001 al 2014

Per il *record linkage* tra i dati di ospedalizzazione e Salute 2000, poiché non si disponeva dei dati identificativi e quindi non era possibile effettuare alcuni controlli, si è deciso di procedere solo per gli individui con chiave completa (126.601).

Per le SDO si è fatto uso del Codice Fiscale presente sull'archivio del Ministero della Salute 2001-2014 (i codici fiscali dei ricoveri non sono disponibili per gli anni 1999-2000). Il *record linkage* è stato fatto in due fasi, la prima dal 2001 al 2008 e la seconda dal 2009 al 2014.

L'ISTAT ha inviato al Ministero della Salute i Codici fiscali degli intervistati corredati di un identificativo che consentisse il riaggancio ai dati dell'indagine e il Ministero ha provveduto al *record linkage* restituendo gli abbinati corredati delle informazioni cliniche relative al ricovero.

Dopo avere acquisito i dati SDO dal Ministero della Salute, l'Istat ha provveduto ad effettuare il *record linkage* con l'indagine sulla salute 1999-2000 e i dati di mortalità dal 1999 al 2014.

4. La costruzione del sistema informativo a partire dall'indagine sulla salute del 2005: nuove strategie e metodi

La procedura applicata all'indagine Istat sulla Salute del 2000 è stata estesa all'edizione 2005 dell'indagine i cui dati sono stati integrati con le informazioni derivanti dai flussi di mortalità e di ospedalizzazione dal 2004 al 2014.

Le strategie adottate presentano delle innovazioni, rispetto a quanto fatto per il sistema integrato costruito a partire dall'indagine sulla salute del 2000.

Anche per l'edizione del 2004-2005 dell'indagine sulla salute, l'Istat ha potuto ampliare la numerosità campionaria da 24.000 a circa 60.000 famiglie grazie ad un contributo del Ministero della Salute e delle Regioni. Il tasso di risposta è dell'83,1% con un campione effettivo di 50.474 famiglie (128.040 rispondenti).

4.1 Acquisizione delle informazioni anagrafiche degli individui intervistati per l'indagine ISTAT Salute 2005 e procedure di ricostruzione e validazione dei codici fiscali

Per creare una base dati longitudinale a partire dall'indagine Salute 2004-2005 era necessario disporre del codice fiscale come chiave per il linkage con gli altri archivi. La novità rispetto all'indagine del 2000 è il fatto che sin dalla fase di estrazione delle famiglie campione, l'Istat ha informatizzato le procedure di gestione della rilevazione (SIGIF Sistema di Gestione delle Indagini sulle Famiglie) e ciò ha consentito di disporre su supporto digitale dei codici fiscali e/o delle informazioni anagrafiche degli individui intervistati.

Le informazioni anagrafiche e i CF acquisiti da SIGIF sono stati controllati e corretti e, in seguito, integrandoli con le informazioni rilevate con l'indagine, è stato ricostruito il codice fiscale. E' stata quindi creata una variabile che descrive l'esito della procedura di calcolo del codice fiscale. Nei casi in cui il codice fiscale era già acquisito da SIGIF è stata fatta una verifica della completezza e della coerenza con i dati anagrafici.

E' stato quindi effettuato un controllo successivo al calcolo del codice fiscale e alla verifica del CF fornito che ha consentito di approfondire i casi di dubbia attribuzione. Infine, sulla base della casistica per tipologia di errore, è stata fatta la scelta tra Codice fiscale fornito e CF calcolato.

4.1.1 *Correzione dei dati identificativi, calcolo del codice fiscale e verifica del codice fiscale fornito*

Una volta acquisiti i dati da SIGIF integrati con quelli derivanti dall'indagine, si è proceduto ad un controllo delle informazioni anagrafiche (nome, cognome, data di nascita, sesso, luogo di nascita e codice fiscale, quando presente) per individuare la presenza e il tipo di eventuali errori formali di registrazione, come ad esempio caratteri non compatibili, campi non compilati, ecc. Sono state fatte quindi delle correzioni.

In particolare, nel caso di *nome e cognome* sono stati accettati, oltre ai caratteri alfabetici, anche gli apici, gli spazi, le virgole, e i trattini (-). Invece sono stati sostituiti:

- gli 0 (zero) con le O;
- i caratteri non ammessi (es. punto * / \ () , “ _), con uno spazio.

Nel caso del *codice fiscale fornito* sono state effettuate le seguenti correzioni:

- la sostituzione degli 0 (zero) con le O e viceversa (a seconda che lo scambio di carattere sia effettuato nelle posizioni del codice fiscale in cui rientrano i caratteri relativi al nome e cognome oppure alla data di nascita o al luogo);
- la sostituzione del carattere X con * per i caratteri in posizione 13,14,15 relativi al luogo;
- il ricalcolo del codice di controllo dove il sedicesimo carattere non è presente.

E' stato quindi calcolato il codice fiscale per tutti i 128.040 intervistati (inclusi gli individui per i quali era già presente un codice fiscale) con l'algoritmo di calcolo predisposto dall'Agenzia delle Entrate. Quando le informazioni anagrafiche non erano complete il codice fiscale è stato calcolato parzialmente. La procedura adottata genera anche una variabile che descrive l'esito del calcolo del codice fiscale le cui modalità sono:

- Codice fiscale non calcolato (mancano dati anagrafici)⁵;
- Codice fiscale calcolato parziale;
- Codice fiscale calcolato completo.

Per quanto riguarda la disponibilità e correttezza del CF dall'archivio SIGIF, si è proceduto ad una verifica della presenza e, quando disponibile, della coerenza con le informazioni anagrafiche. Per descrivere l'esito di questa verifica e fornire un dettaglio delle eventuali incoerenze riscontrate sono state predisposte due variabili. La prima descrive l'esito della verifica e può assumere le seguenti modalità:

1. Codice fiscale assente;
2. Codice fiscale con formato errato (anche per omocodia);
3. Codice fiscale non coerente con le informazioni anagrafiche;
4. Codice fiscale corretto (verifica parziale con possibile omocodia);
5. Codice fiscale corretto (verifica parziale);
6. Codice fiscale corretto (possibile omocodia);
7. Codice fiscale corretto.

E' stata quindi creata una seconda variabile che descrive il dettaglio dell'eventuale errore riscontrato sia nel calcolo del codice fiscale (cioè quando il codice risulta "calcolato parziale", oppure "non calcolato") sia nel codice fornito per i casi in cui il codice fiscale risulta con formato errato (punto 2) o non coerente con le informazioni anagrafiche (punto 3).

La variabile combina i possibili errori presenti nei diversi campi che compongono il CF. L'errore può fare riferimento alle seguenti informazioni anagrafiche:

COG - Cognome
 NOM - Nome
 YEA - Anno nascita
 MON - Mese nascita
 DAY - Giorno nascita
 SEX - Sesso
 LUO - Luogo di nascita (Italia)
 EST - Luogo di nascita (Esteri)
 CON - Carattere di controllo

Ognuna di queste informazioni può ricadere nei seguenti casi:

V –Vuoto (l'informazione è mancante)
 E – Errato (l'informazione è errata)
 C – Corretto automaticamente (sono state applicate delle correzioni - solo per Cognome e Nome)

La variabile di dettaglio dell'errore può assumere delle modalità che sono combinazioni delle due variabili sopra descritte. Nella tabella di seguito vengono forniti alcuni esempi.

Tavola 5 – Esempi di errori nei codici fiscali calcolati con esito 'Codice fiscale calcolato parziale' - Salute 2004-2005

Dettaglio del calcolo del CF	Informazioni mancanti
COG:V;NOM: V	Cognome e Nome non disponibili
MON: V; LUO: V;	Mese e luogo di nascita non disponibili
COG: V; NOM: V; LUO: V; EST: V	Cognome, Nome e luogo o Stato Estero di nascita non disponibili

⁵ Il Codice fiscale non calcolato non si verifica nella casistica degli esiti relativi al collettivo degli intervistati in occasione dell'indagine sulla Salute condotta nel 2004-2005 in quanto le informazioni presenti nel Sistema di Gestione delle Indagini sulle Famiglie sono state integrate con quelle rilevate con l'indagine ed è sempre presente la data di nascita e il sesso degli intervistati.

4.1.2 I controlli e la scelta del CF per il linkage con mortalità e ospedalizzazione

Le procedure di calcolo del codice fiscale e di verifica del codice fiscale fornito (da SIGIF) sono state completate con dei controlli che sono stati effettuati per i casi i cui codici fiscali forniti non risultavano coerenti con l'anagrafica e il CF calcolato era formalmente corretto ma diverso dal CF fornito e non era quindi accertabile se l'errore fosse nel CF fornito o nell'anagrafica.

La Tavola 6 documenta il numero e la completezza dei CF presenti nell'archivio SIGIF (indicati come CF forniti) e dei CF calcolati partendo dai dati anagrafici SIGIF.

I codici fiscali già presenti nell'archivio sono 92.304 (circa il 72 per cento del totale dei record). Dei 128.040 codici ricostruiti, 125.425 sono formalmente validi (98 per cento), cioè non presentano parti mancanti, 2.615 (2 per cento) risultano parziali per mancanza di alcune delle informazioni anagrafiche (ad esempio in alcuni casi non erano presenti il nome, il cognome, il luogo di nascita, ecc.).

Tavola 6 – Distribuzione dei codici fiscali forniti e calcolati per livello di completezza - Salute 2004-2005

Esito del calcolo del CF e della verifica del CF fornito	CF fornito		CF calcolato	
	v.a.	%	v.a.	%
Assente	35.736	27,9	-	0,0
Parziale	36	0,0	2.615	2,0
Completo	92.268	72,1	125.425	98,0
TOTALE	128.040	100,0	128.040	100,0

Il dettaglio relativo alle parti mancanti dei codici fiscali calcolati è riportato nella Tavola 7. In 2.188 casi le informazioni anagrafiche sono totalmente mancanti, fatta eccezione per la data di nascita e il sesso che sono rilevati con l'intervista. In 427 manca l'informazione sul luogo di nascita; per 76 di questi record il CF è stato ricostruito sostituendo alle informazioni relative al luogo di nascita quelle relative al luogo di residenza.

Tavola 7 – Codici fiscali calcolati per completezza e dettaglio delle parti mancanti - Salute 2004-2005

Esito del CF calcolato	CF calcolato	Parti mancanti	v.a.	%
Completo			125.425	97,96
Parziale			2.615	2,04
	di cui			
	ABCDEF YYMDD XXXXX	Luogo di nascita/Stato estero e Codice controllo	427	0,33
	XXXXXXXXYYMDDXXXXX	Cognome, Nome, Luogo di nascita/Stato estero e Codice controllo	2.188	1,71
TOTALE			128.040	100,00

Confrontando l'esito della verifica effettuata su entrambi i CF (Tavola 8) si osserva che per 87.807 (68,6%) individui il CF fornito e quello calcolato sono completi ed uguali. In questo caso, la scelta del CF da utilizzare come chiave di aggancio per il successivo *record linkage* non comporta problemi.

Per i 35.736 individui per i quali il CF fornito non era disponibile, il CF calcolato è risultato completo e formalmente corretto in 33.472 casi (93,7%) mentre nei restanti 2.264 casi il CF calcolato è solo parziale (di questi 2.188 non sono recuperabili come mostrato nella Tavola 7).

Sono stati verificati separatamente i 17 casi in cui il CF fornito risulta "corretto con possibile omocodia". A tal fine, il CF è stato ricostruito secondo le disposizioni dell'Agenzia delle Entrate e i codici così ottenuti sono risultati identici a quelli forniti che perciò si possono considerare formalmente corretti.

Tavola 8 – Esiti della verifica effettuata sui codici fiscali forniti e calcolati - Salute 2004-2005

Esito CF fornito	Esito CF calcolato		Totale
	CF completo	CF parziale	
CF fornito uguale a CF calcolato	87.807		87.807
CF fornito non presente	33.472	2.264	35.736
CF fornito con formato errato (anche per omocodia)	208		208
CF fornito corretto (possibile omocodia)	17		17
CF fornito corretto (verifica parziale)	23	347	370
CF fornito non coerente con dati individuo	3.898	4	3.902
Totale	125.425	2.615	128.040

Dei 370 CF forniti il cui esito della verifica era “CF fornito corretto con verifica parziale”, 23 sono parziali ma il CF calcolato è completo e formalmente corretto perché sono disponibili tutte le informazioni necessarie. I restanti 347 CF forniti sono completi ma è stato possibile fare solo una verifica parziale perché le informazioni anagrafiche non erano complete e quindi il CF calcolato risulta parziale.

La valutazione dei 3.902 record (circa il 3 per cento del totale) per i quali il CF fornito risulta non coerente con i dati dell’individuo ha richiesto degli approfondimenti. Preliminarmente questi codici sono stati classificati per tipologia di incongruenza riscontrata della quale si fornisce il dettaglio nella Tavola 9. La maggior parte dei CF forniti differisce dai CF calcolati a causa di difformità nella parte relativa al nome o al luogo di nascita. Nei casi in cui la difformità tra i due codici fiscali è riconducibile alla parte relativa al nome viene verificato che l’individuo corrispondente abbia più nomi. In caso di individui con nomi multipli, il CF fornito viene confrontato con un nuovo CF generato facendo riferimento solo al primo nome dell’individuo. I CF forniti che risultano uguali ai CF calcolati riferendosi solo al primo nome sono considerati formalmente corretti. Nel caso dei CF forniti che differiscono dai CF calcolati nella parte che si riferisce al luogo di nascita (che rappresentano quasi il 50% di questi casi) viene verificato che il luogo corrispondente abbia subito una variazione di codice catastale. Nei casi in cui si accerti la variazione catastale, il CF fornito è ritenuto formalmente corretto.

Tutti i casi rimanenti sono stati sottoposti a dei controlli predisposti in base alle diverse tipologie di errore rilevato. In particolare, prima della valutazione di coerenza formale dei codici, è stata messa in atto una procedura che ha permesso di confrontare il codice fiscale fornito e il codice fiscale calcolato attraverso una misura della “somialianza”. A tal fine, si è fatto ricorso a una funzione Sas “Spedis” che ha permesso di generare una variabile che fornisce la “distanza” tra il codice fiscale fornito e il codice fiscale calcolato.

La verifica dei casi più problematici è stata effettuata dopo aver esaminato i codici fiscali singolarmente.

Tavola 9 – Distribuzione dei CF forniti non coerenti con le informazioni anagrafiche per tipologia di errore - Salute 2004-2005

Tipologia errore	v.a.	%
Luogo di nascita	1.877	48,1
Nome	811	20,8
Carattere di controllo	346	8,9
Errori Multipli	278	7,0
Cognome	272	7,0
Giorno di nascita	92	2,4
Anno di nascita	77	2,0
Sesso	75	1,9
Mese di nascita	74	1,9
TOTALE	3.902	100,0

La valutazione dei codici fiscali secondo le procedure di analisi illustrate nel testo, ha permesso di scegliere i codici fiscali da utilizzare come chiave di linkage a cui abbinare le informazioni sui decessi e i ricoveri. I codici scelti sono distribuiti tra codici fiscali forniti e calcolati come riportato nella Tavola 10.

Il 98,3% dei CF forniti o calcolati è utilizzabile per il linkage.

Dei 128.040 intervistati non sono abbinabili solo 2.188 casi per i quali il codice fiscale fornito era assente e quello calcolato parziale in quanto per questi individui erano presenti solo le informazioni rilevate con l'indagine (data di nascita e sesso).

Inoltre non sono abbinabili 2 casi con codice fiscale fornito e calcolato uguali a causa di una duplicazione dell'anagrafica verosimilmente imputabile ad errori di registrazione (si passa quindi da 87.807 casi a 87.805 casi).

Tavola 10 – Distribuzione dei codici fiscali forniti e calcolati per esito e tipo di codice scelto per il linkage - Salute 2004-2005

Esito della verifica del CF calcolato (CF-1)	Esito della verifica del CF fornito nell'indagine Salute 2005 (CF-2)	Intervistati	Abbinabili	CF scelto per il linkage
		v.a.	v.a.	
Completo	Completo coerente	87.807	87.805	CF-2
Completo	Formato errato	208	208	CF-1
Completo	Completo coerente con possibile omocodia	17	17	CF-2
Completo	Non coerente	3.811	3.811	CF-1
Completo	Non coerente*	87	87	CF-2
Completo	Parziale	23	23	CF-1
Completo	Assente	33.472	33.472	CF-1
Parziale	Non coerente	4	4	CF-2
Parziale	Assente	2.264	76	CF-1
Parziale	Completo verifica parziale	347	347	CF-2
Totale		128040	125850	

* Con codici catastali cambiati e doppi nomi

4.2 Il *record linkage* tra indagine salute e archivio mortalità: la creazione dei codici fiscali dei decessi e le nuove procedure

La procedura impiegata per la ricostruzione, l'analisi, e la validazione dei CF degli individui contenuti nel campione dell'indagine Salute 2004-2005, nonché quella per la scelta del CF (tra fornito e calcolato) da utilizzare per il linkage, è stata applicata anche ai dati relativi ai decessi registrati tra il 2004 e il 2012.

A partire dal 2013 c'è un ulteriore cambiamento e i dati relativi ai decessi si acquisiscono direttamente corredati della variabile di scelta del CF dalla procedura implementata e standardizzata per il progetto Istat "Differenze socio-economiche nella mortalità", nell'ambito del quale sono state integrate, mediante record-linkage, le informazioni del Censimento della popolazione e delle abitazioni del 2011 con quelle dell'indagine sui decessi e le cause di morte per gli anni 2012, 2013 e 2014⁶.

Il *record linkage* tra Salute 2005 e mortalità è stato effettuato dall'ISTAT con una metodologia ottimizzata grazie alla esperienza maturata nelle prime fasi sperimentali del progetto.

⁶ Il progetto nasce dall'esigenza di misurare i differenziali sociali nella mortalità, non solo a livello nazionale, ma anche regionale e provinciale avvalendosi della migliore fonte informativa a livello nazionale per le variabili di posizione sociale, vale a dire il Censimento della Popolazione. La *performance* del record linkage è molto alta: il 97% dei decessi eleggibili.

4.2.1 La creazione e la scelta del CF dei decessi

Come per l'indagine Salute 2004-2005 si è proceduto ad una verifica dei codici fiscali presenti nell'archivio ISTAT e ad una selezione dei CF abbinabili. L'esito di queste procedure per i decessi dal 2004 al 2012 è descritto nella Tavola 11 e nella Tavola 12.

Tavola 11 – Decessi della popolazione italiana per esito della verifica del Codice fiscale fornito dal 2004 al 2012

Esito della verifica del CF fornito	2004	2005	2006	2007	2008	2009	2010	2011	2012	TOTALE
Completo coerente uguale a CF calcolato	366.495	418.088	422.695	443.325	456.891	456.201	457.398	469.682	493.196	3.983.971
Assente	93.402	89.674	80.317	75.597	70.668	68.598	66.840	69.387	68.489	682.972
Formato errato (anche per omocodia)	1.025	296	216	214	224	638	1.014	1.136	1.132	5.895
Completo coerente (possibile omocodia)	4	8	8	12	16	26	52	52	44	222
Completo coerente verifica parziale (anche con possibile omocodia)	1	0	0	1	0	0	1	1	0	4
Completo coerente verifica parziale	13.979	13.724	14.121	15.022	15.625	17.154	20.557	22.387	24.343	156.912
Non coerente	70.145	46.538	41.257	38.710	38.046	45.820	38.753	31.508	26.316	377.093
TOTALE	545.051	568.328	558.614	572.881	581.470	588.437	584.615	594.153	613.520	5.207.069

Come si può osservare nella Tavola 12, per il 99,5% dei decessi è possibile effettuare il linkage con altre fonti utilizzando il CF presente nell'archivio di mortalità o quello ricostruito a partire dalle informazioni anagrafiche disponibili.

Tavola 12 – Decessi abbinabili per esito della verifica del CF fornito e percentuale degli abbinabili dal 2004 al 2012

Esito della verifica del CF fornito	2004	2005	2006	2007	2008	2009	2010	2011	2012	TOTALE
Completo coerente uguale a CF calcolato	366.495	418.088	422.695	443.325	456.891	456.201	457.398	469.682	493.196	3.983.971
Assente	88.630	85.375	78.377	73.739	68.986	66.525	64.751	67.086	66.255	659.724
Formato errato (anche per omocodia)	964	242	162	177	187	586	961	1.077	1.044	5.400
Completo coerente (possibile omocodia)	4	8	8	12	16	26	52	52	44	222
Completo coerente verifica parziale (anche con possibile omocodia)	1	0	0	1	0	0	1	1	0	4
Completo coerente verifica parziale	13.971	13.720	14.116	15.020	15.618	17.143	20.538	22.360	24.316	156.802
Non coerente	70.141	46.535	41.257	38.707	38.045	45.818	38.749	31.504	26.311	377.067
TOTALE	540.206	563.968	556.615	570.981	579.743	586.299	582.450	591.762	611.166	5.183.190
Percentuale degli abbinabili sui decessi	99,1	99,2	99,6	99,7	99,7	99,6	99,6	99,6	99,6	99,5

La qualità del CF riportato nel certificato di morte può considerarsi buona dato che il CF fornito e il CF calcolato sono uguali nel 76,5% dei casi (Tavola 13). Per il 13,1% dei decessi il CF non è disponibile negli archivi di mortalità ma la sua ricostruzione, utilizzando le informazioni anagrafiche presenti nell'archivio, consente un recupero quasi totale (12,7%).

Dal 2013 i dati relativi ai decessi vengono acquisiti già corredati della variabile di scelta del Codice fiscale dall'archivio del progetto ISTAT "Differenze socio-economiche nella mortalità" per il quale sono state applicate le procedure di calcolo e validazione dei CF dei decessi già sperimentate in questo studio⁷. Nella Tavola 14, che si riferisce ai decessi dal 2013 al 2014, la quota degli abbi-

⁷ Poiché nell'ambito di questo progetto si effettua il *record linkage* dei decessi con la popolazione al Censimento 2011, le procedure sono state ottimizzate e standardizzate per consentire l'aggiornamento annuale del database.

nabili con altre fonti è pari al 99,7%. Il CF fornito e il CF calcolato sono uguali nell'84,2% dei casi, per 127.976 decessi (10,7% del totale) il CF non è disponibile negli archivi di mortalità ma la copertura è pressoché totale utilizzando il CF calcolato (124.921 decessi).

Tavola 13 – Performance dell'utilizzabilità per il linkage dei CF dei decessi 2004-2012

Esito della verifica del CF fornito	Decessi 2004-2012	Abbinabili con CF fornito o calcolato 2004-2012	Percentuale degli abbinabili sul totale dei decessi
Completo coerente uguale a CF calcolato	3.983.971	3.983.971	76,5
Assente	682.972	659.724	12,7
Formato errato	5.895	5.400	0,1
Completo coerente (possibile omocodia)	222	222	0,0
Completo coerente verifica parziale (anche con possibile omocodia)	156.916	156.806	3,0
Non coerente	377.093	377.067	7,2
TOTALE	5.207.069	5.183.190	99,5

Tavola 14 – Distribuzione dei codici fiscali forniti e calcolati per esito e tipo di codice scelto per il linkage, decessi 2013 e 2014

Esito della verifica del CF calcolato (CF-1)	Esito della verifica del CF fornito (CF-2)	Decessi	Abbinabili	CF scelto per il linkage
		v.a.	v.a.	
Completo	Completo coerente	981.926	981.926	CF-2
Completo	Formato errato	1.965	1.965	CF-1
Completo	Non coerente	40.010	40.010	CF-1
Completo	Parziale	1.018	1.018	CF-1
Completo	Assente	124.921	124.921	CF-1
Parziale	Formato errato	123	0	Nessuno
Parziale	Non coerente	1.500	1.500	CF-2
Parziale	Parziale	46	0	Nessuno
Parziale	Assente	3.055	0	Nessuno
Parziale	Completo verifica parziale	36.677	36.677	CF-2
Totale		1.191.241	1.188.017	

4.2.2 Il record linkage tra l'indagine Salute 2005 e i decessi 2004-2014

Portata a termine la procedura di selezione dei codici fiscali degli intervistati nell'ambito dell'indagine sulla salute condotta nel 2004-2005 e delle persone decedute dal 2004 al 2014, sono state avviate le procedure di *record linkage*. L'esito di queste procedure ha portato alla creazione di tre distinti dataset, il primo che include gli individui per i quali è stato fatto il *record linkage* con CF uguale e completo nei due archivi, il secondo che contiene i casi di individui intervistati nell'indagine Salute 2005 per i quali non è risultato nessun aggancio con i decessi e un terzo contenente i deceduti per i quali non c'è stata corrispondenza nel campione dell'indagine salute.

Per questi due ultimi dataset, con i passi successivi della procedura, sono stati resi parziali i CF di entrambi gli archivi eliminando la parte corrispondente al luogo e al codice di controllo (5 digit del CF) ed è stato effettuato un altro passo di linkage tra i decessi e Salute 2005 non abbinati al primo tentativo con i CF resi parziali. E' stato poi effettuato un controllo degli abbinati con CF parziale per valutare in quali casi il *record linkage* fosse accettabile. A questo scopo è stata utilizzata una funzione del software Sas (funzione "Spedis") con la quale sono state analizzate e misurate le differenze tra le informazioni presenti nei due archivi relative al cognome e al luogo di nascita ed è

stata verificata inoltre la presenza di duplicati generati dall'utilizzo di CF parziali.

Sempre sui due gruppi dei non abbinati, quello degli intervistati non associati a decessi e quello dei deceduti senza corrispondenza tra gli intervistati, è stato fatto un ulteriore passo di linkage con un CF reso parziale nel quale è stata eliminata la parte corrispondente al nome. Anche per questo gruppo di casi abbinati si è effettuato un controllo con Sas “Spedis”, in questo caso delle differenze nelle informazioni relative al nome e al cognome.

Per entrambi i controlli di “distanza” effettuati con Sas “Spedis” è stata definita una soglia di accettazione sulla base dei risultati empirici validata anche attraverso controlli ad hoc sui casi più incerti.

Al termine di queste procedure è stato possibile effettuare il linkage tra Salute 2005 e mortalità per i seguenti casi:

- Linkage con CF di salute e mortalità completamente identici
- Linkage con CF identici ad eccezione del luogo (luogo assente o in Salute o in Mortalità)
- Linkage con CF identici ad eccezione del luogo (distanza luogo minore di una certa soglia)
- Linkage con CF identici ad eccezione del nome

È stato quindi creato un file dati intermedio contenente gli identificativi di Salute 2005 e dei decessi dal 2004 al 2014 utilizzato per il successivo aggancio con le cause di morte e le altre variabili dell'archivio mortalità di interesse per lo studio longitudinale.

Complessivamente della coorte dei 125.850 intervistati nell'ambito dell'indagine sulla salute del 2005 e abbinabili (perché forniti delle informazioni necessarie a costruire la chiave di linkage), risultano deceduti 11.050 individui.

4.3 Il *record linkage* con i dati di ospedalizzazione del Ministero della Salute dal 2004 al 2014

Per il *record linkage* tra i dati di ospedalizzazione e Salute 2005, si è fatto uso del Codice Fiscale già presente sull'archivio SDO 2004-2014.

L'ISTAT ha inviato al Ministero della Salute i Codici fiscali degli intervistati corredati di un identificativo che consentisse il riaggancio ai dati dell'indagine e il Ministero ha provveduto al *record linkage* restituendo gli abbinati corredati delle informazioni cliniche relative al ricovero.

Dopo avere acquisito i dati SDO dal Ministero della Salute, l'Istat ha provveduto ad effettuare il *record linkage* con l'indagine sulla salute 2004-2005 e i dati di mortalità dal 2004 al 2014.

4.4 L'aggiornamento del Follow-up 2005 con i trasferimenti all'estero: il *record linkage* con i dati ISCAN

Il database relativo all'indagine sulla Salute condotta nel 2004-2005 è stato integrato con i dati dal 2005 al 2014 provenienti dalla rilevazione Istat “Iscrizioni e cancellazioni all'anagrafe per trasferimento di residenza” (ISCAN) utilizzando il CF come chiave identificativa univoca per il *record linkage*.

Il CF utilizzato per la popolazione e per i trasferimenti è quello disponibile nell'Anagrafe Virtuale Statistica (ANVIS)⁸.

Sono state selezionate esclusivamente le informazioni relative ai trasferimenti verso l'estero, escludendo i rientri e gli eventuali altri trasferimenti. In particolare sono stati agganciati i dati relativi a mese e anno di trasferimento. Ciò consente di escludere dalla popolazione del follow-up questi individui nel momento in cui non fanno più parte del collettivo.

La numerosità dei trasferimenti è molto bassa ma è utile quantificarla per le analisi anche future sulla popolazione straniera (Tavola 15).

⁸ L'Anagrafe Virtuale Statistica (ANVIS) assume come popolazione di riferimento la popolazione residente iscritta in anagrafe. L'archivio è costruito a partire dai microdati della Popolazione legale alla data del censimento del 9.10.2011 considerati i ritorni di SIREA (Sistema individuale di revisione anagrafica) ed è alimentato in modo continuo dai microdati di flusso degli eventi della dinamica demografica (nascite, decessi, migrazioni interne, migrazioni internazionali, iscrizioni e cancellazioni per altri motivi, acquisizioni di cittadinanza, altre variazioni o mutazioni). Il sistema di gestione ANVIS consente la storicizzazione degli eventi di dinamica demografica a partire dal censimento.

Tavola 15 – Individui trasferiti all'estero per anno trasferimento - Follow-up 2005

Anno di trasferimento	v.a.	%
2005	6	1,2
2006	36	7,4
2007	30	6,2
2008	30	6,2
2009	43	8,9
2010	44	9,1
2011	48	9,9
2012	84	17,4
2013	68	14,1
2014	95	19,6
Totale	484	100,0

5. Il trattamento e il controllo di qualità dei dati

5.1 Check e trattamento dei dati

Per eliminare incoerenze ed errori, sono stati effettuati controlli e correzioni sui singoli archivi e, al termine delle singole fasi delle operazioni di *record linkage*, sui database nei quali le tre fonti sono state integrate (Follow-up 2000 e Follow-up 2005).

Per quanto riguarda le dimissioni ospedaliere (SDO) linkate con le indagini sulla salute, sono stati fatti dei controlli sulla completezza e correttezza delle singole variabili successivamente rilasciate oppure utilizzate per la verifica di coerenza delle informazioni all'interno di ciascun archivio e della congruenza tra i tre archivi. Si è quindi proceduto a verificare le coerenze tra le variabili SDO.

Particolare attenzione è stata dedicata alle variabili relative ai dati anagrafici, all'anno di dimissione e di decesso, all'età di ricovero e di morte e, per quanto riguarda il Follow-up 2005, alla cittadinanza, rilevante per le analisi sulla popolazione straniera.

Per il solo Follow-up 2000 il *record linkage* con le SDO è avvenuto in due fasi (dal 2001 al 2008 e dal 2009 al 2014). Nella seconda fase si è reso necessario un lavoro di riallineamento delle variabili SDO per le modifiche nel frattempo avvenute, ad esempio nelle modalità o nei formati.

Si è quindi proceduto ad effettuare il controllo di coerenza tra i tre archivi e il trattamento dei dati finalizzati a verificare la presenza di eventuali errori nel linkage o ad eliminare informazioni incoerenti.

I principali controlli effettuati hanno riguardato la coerenza tra:

- età alla morte dell'archivio mortalità e età alla morte calcolata a partire dall'età all'intervista;
- date di nascita dall'archivio di mortalità e dell'indagine salute;
- modalità di dimissione SDO=1 (decesso) e presenza di ricoveri successivi;
- modalità di dimissione SDO=1 (decesso) e presenza di record mortalità;
- data di decesso con data di ricovero e di dimissione;
- regime di ricovero e giorni di ricovero.

Sono stati effettuati numerosi altri controlli ma l'entità delle correzioni apportate è comunque molto contenuta.

Al termine di questi controlli i due database aggiornati al 2014 presentano le numerosità illustrate nella Tavola 16.

Tavola 16 – Quadro sinottico Follow-up 2000 e Follow-up 2005

	Followup 2000	Followup 2005
Individui intervistati	140.011	128.040
Individui abbinabili con decessi	128.818	125.850
Individui abbinabili con SDO	126.601	125.850
Decessi	17.508	11.050
SDO	282.648	188.151
Individui con almeno 1 ricovero	80.182	66.061

5.2. Controllo potenziale bias per copertura parziale del campione

Il lavoro di ricostruzione e selezione dei codici fiscali degli intervistati nell'ambito delle indagini Salute 2000 e Salute 2005 ha consentito di raggiungere un buon livello di copertura del campione effettivo. Per l'indagine Salute 2000 dei 140.011 individui intervistati, 128.818 individui sono risultati corredati delle informazioni necessarie per il linkage con la mortalità (92,1%) e 126.601 individui (quelli per i quali si dispone di un CF completo) con i ricoveri (90,4%). Per l'indagine Salute 2005 i risultati migliorano. Infatti dei 128.040 individui presenti nel campione, il 98,3%, pari a 125.850 persone risulta "linkabile" con entrambe le fonti.

Per valutare le eventuali distorsioni dovute ad una parziale copertura del campione in entrambi i dataset è stata effettuata la comparazione per profilo tra gli inclusi nel follow-up e gli esclusi per mancanza di chiave di linkage.

Utilizzando i pesi di riporto all'universo sono stati effettuati confronti per sesso e classi di età, per titolo di studio e condizione lavorativa, per alcuni indicatori di salute (percezione dello stato di salute e disabilità) e per i fattori di rischio relativi all'abitudine al fumo e al BMI⁹.

Per quanto riguarda il database costruito dall'indagine Salute 2000, analizzando la distribuzione per classi di età e genere (Tavola 17) emergono alcune differenze statisticamente significative fra non abbinabili e abbinabili per i minori di 14 anni, soprattutto per i maschi, e nelle fasce di età 35-44 e 55-64. Si osservano differenze più lievi per le donne tra i 45 e i 54 anni e tra quelle di 75 anni e più. Tali differenze tuttavia non attenuano la rappresentatività degli abbinabili che, per tutte le classi di età considerate e per genere, sono sempre allineati con il totale del campione. L'analisi condotta con il test del χ^2 conferma quanto già osservato: non ci sono differenze statisticamente significative tra i due collettivi per genere mentre la distribuzione si differenzia per età anche se in misura contenuta.

⁹ Il Body Mass Index (BMI) è dato dal rapporto tra il peso corporeo espresso in chilogrammi e il quadrato dell'altezza espressa in metri. Una volta calcolato il BMI per la classificazione dell'eccesso di peso, un valore dell'indice inferiore a 18,5 identifica le persone sottopeso, maggiore o uguale di 18,5 ma inferiore a 25 le persone normopeso. Un BMI maggiore o uguale a 25 ma inferiore a 30 indica il sovrappeso e infine le persone sono considerate obese se il valore del BMI è maggiore o uguale a 30.

Tavola 17 – Follow-up 2000 - Abbinabili e non abbinabili per sesso e classe di età (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Maschi								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%		%	IC95%		%	IC95%	
0-14	19,4	18,0	20,8	15,1	14,7	15,5	15,4	15,0	15,9
15-24	12,9	11,7	14,1	12,3	12,0	12,7	12,4	12,0	12,7
25-34	16,9	15,6	18,2	16,0	15,6	16,4	16,1	15,6	16,5
35-44	13,9	12,7	15,2	16,1	15,7	16,5	15,9	15,5	16,3
45-54	12,4	11,2	13,6	13,5	13,2	13,9	13,4	13,0	13,8
55-64	9,8	8,8	10,9	12,0	11,6	12,4	11,8	11,4	12,2
65-69	4,8	4,1	5,5	5,1	4,8	5,3	5,1	4,8	5,3
70-74	3,5	2,8	4,1	4,3	4,1	4,6	4,3	4,1	4,5
75-79	3,4	2,8	4,0	3,2	3,0	3,4	3,2	3,0	3,4
>=80	3,0	2,4	3,6	2,4	2,2	2,5	2,4	2,3	2,6
Totale	100,0			100,0			100,0		

Classe di età	Femmine								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%		%	IC95%		%	IC95%	
0-14	16,0	14,7	17,2	13,5	13,1	13,9	13,7	13,3	14,1
15-24	12,0	10,9	13,1	11,3	11,3	11,4	11,4	11,0	11,7
25-34	14,7	13,5	15,9	14,8	14,3	15,2	14,8	14,4	15,2
35-44	13,3	12,2	14,5	15,2	14,8	15,7	15,1	14,7	15,5
45-54	11,6	10,5	12,7	13,2	12,8	13,6	13,1	12,7	13,5
55-64	10,3	9,3	11,4	11,9	11,5	12,3	11,8	11,4	12,1
65-69	5,6	4,8	6,4	5,8	5,6	6,1	5,8	5,6	6,1
70-74	4,8	4,1	5,5	5,1	4,8	5,3	5,0	4,8	5,3
75-79	5,9	5,1	6,7	4,5	4,3	4,7	4,6	4,4	4,9
>=80	5,9	5,1	6,7	4,6	4,4	4,9	4,7	4,5	4,9
Totale	100,0			100,0			100,0		

Classe di età	Totale								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%		%	IC95%		%	IC95%	
0-14	17,6	16,6	18,6	14,3	14,0	14,6	14,6	14,3	14,9
15-24	12,4	11,6	13,2	11,8	11,6	12,1	11,9	11,6	12,1
25-34	15,7	14,8	16,7	15,4	15,1	15,7	15,4	15,1	15,7
35-44	13,6	12,8	14,5	15,6	15,4	15,9	15,5	15,2	15,8
45-54	12,0	11,2	12,8	13,4	13,1	13,6	13,3	13,0	13,5
55-64	10,1	9,4	10,9	11,9	11,7	12,2	11,8	11,5	12,0
65-69	5,2	4,7	5,7	5,5	5,3	5,6	5,4	5,3	5,6
70-74	4,2	3,7	4,7	4,7	4,5	4,9	4,7	4,5	4,8
75-79	4,7	4,2	5,2	3,9	3,7	4,0	4,0	3,8	4,1
>=80	4,5	4,0	5,0	3,5	3,4	3,7	3,6	3,5	3,7
Totale	100,0			100,0			100,0		

La quota del campione che non può essere integrata con i dati di mortalità e ospedalizzazione non è diversa dall'intera popolazione e dagli abbinabili per quanto riguarda il titolo di studio controllato suddividendo la popolazione in adulta (25-64 anni) e anziana (65 anni e più) (Tavola 18). Ciò è confermato dall'ulteriore analisi condotta con un test del χ^2 sulle proporzioni.

Tavola 18 – Follow-up 2000 - Abbinabili e non abbinabili per classe di età e titolo di studio (valori percentuali e intervalli di confidenza al 95%)

Titolo di studio	Classe di età								
	25<=64								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%
Laurea	10,3	9,3	11,4	9,6	9,3	9,9	9,7	9,4	10,0
Diploma superiore	37,5	35,8	39,2	35,2	34,7	35,7	35,4	34,9	35,8
Licenza media	30,5	28,9	32,1	32,1	31,7	32,6	32,0	31,5	32,5
Licenza elementare	21,7	20,3	23,2	23,1	22,6	23,5	23,0	22,5	23,4
Totale	100,0			100,0			100,0		

Titolo di studio	Classe di età								
	65 anni e più								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%
Laurea	3,5	2,5	4,6	3,0	2,7	3,3	3,1	2,8	3,4
Diploma superiore	7,9	6,4	9,4	8,8	8,3	9,3	8,7	8,2	9,2
Licenza media	11,5	9,7	13,3	11,9	11,3	12,5	11,9	11,3	12,4
Licenza elementare	77,1	74,6	79,5	76,2	75,4	77,0	76,3	75,5	77,1
Totale	100,0			100,0			100,0		

Titolo di studio	Totale								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%
Laurea	8,5	7,7	9,3	8,0	7,8	8,3	8,1	7,8	8,3
Diploma superiore	29,7	28,3	31,0	28,9	28,5	29,3	29,0	28,6	29,4
Licenza media	25,4	24,1	26,8	27,3	26,9	27,7	27,2	26,8	27,6
Licenza elementare	36,4	34,9	37,8	35,7	35,3	36,2	35,8	35,3	36,2
Totale	100,0			100,0			100,0		

Analizzando la distribuzione geografica si osservano invece lievi differenze statisticamente significative (confermate dal test del χ^2) con quote più alte di non abbinabili per il Sud tra gli anziani e per le isole nella popolazione adulta (Tavola 19). La quota dei non abbinabili è invece più bassa degli abbinabili tra gli anziani residenti nel Nord Ovest. E' da considerare, tuttavia, che anche in questo caso la distribuzione degli abbinabili non si discosta da quella della popolazione.

Tavola 19 – Follow-up 2000 - Abbinabili e non abbinabili per classe di età e ripartizione geografica (valori percentuali e intervalli di confidenza al 95%)

Ripartizione geografica	Classe di età								
	<=64								
	Non abbinabili		Abbinabili		Totale				
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	
Nord-ovest	22,9	18,8	26,9	26,1	25,7	26,5	25,8	25,5	26,2
Nord-est	19,8	16,6	23,1	17,8	17,6	18,1	18,0	17,7	18,3
Centro	18,3	15,1	21,5	18,9	18,6	19,2	18,8	18,6	19,1
Sud	23,3	19,9	26,7	25,5	25,1	25,9	25,3	24,9	25,7
Isole	15,7	12,5	18,8	11,7	11,4	12,0	12,0	11,7	12,3
Totale	100,0		100,0			100,0			

Ripartizione geografica	Classe di età								
	65 anni e più								
	Non abbinabili		Abbinabili		Totale				
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	
Nord-ovest	20,1	17,5	22,7	28,2	27,3	29,1	27,5	26,7	28,3
Nord-est	21,0	18,8	23,2	19,8	19,2	20,5	19,9	19,3	20,6
Centro	21,0	18,8	23,2	21,0	20,3	21,6	21,0	20,3	21,6
Sud	25,8	23,4	28,2	20,7	19,9	21,5	21,2	20,4	21,9
Isole	12,2	10,2	14,0	10,3	9,7	10,9	10,4	9,9	11,0
Totale	100,0		100,0			100,0			

Ripartizione geografica	Totale								
	Non abbinabili		Abbinabili		Totale				
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	
Nord-ovest	22,3	21,2	23,5	26,5	26,1	26,8	26,1	25,7	26,5
Nord-est	20,1	19,1	21,0	18,2	17,9	18,4	18,3	18,1	18,6
Centro	18,8	17,9	19,7	19,3	19,0	19,5	19,2	18,9	19,5
Sud	23,8	22,7	24,9	24,7	24,3	25,0	24,6	24,2	25,0
Isole	15,0	14,1	16,0	11,4	11,1	11,7	11,7	11,4	12,0
Totale	100,0		100,0			100,0			

Positiva la *performance* del *record linkage* per quanto riguarda la condizione economica e quella lavorativa. Per entrambe le dimensioni, esaminate per classi di età, il profilo dei non abbinabili non si discosta da quello degli abbinabili e dell'intera popolazione. Osservando i principali indicatori di salute, mentre per la salute percepita non si rilevano eterogeneità, per la disabilità si evidenzia una maggiore prevalenza di non abbinabili tra le donne e conseguentemente nel totale della popolazione ma queste differenze non si osservano confrontando gli abbinabili con il totale del campione. Per quanto riguarda l'abitudine al fumo e l'eccesso di peso l'analisi mostra una soddisfacente corrispondenza della porzione di campione non integrabile con le altre fonti rispetto al totale della popolazione e alla quota degli abbinabili (Appendice A). Lievi scostamenti si osservano tra quanti non hanno mai fumato (soprattutto tra gli anziani) e tra coloro che hanno fumato in passato (con quote più basse tra i non abbinabili rispetto agli abbinabili) e queste differenze sono confermate dal test del χ^2 .

Per il *record linkage* effettuato a partire dall'indagine Salute 2005, complessivamente i risultati sono soddisfacenti. Solo in pochi casi si osservano infatti differenze statisticamente significative tra la distribuzione degli abbinabili e quella relativa all'intera popolazione per tutte le variabili considerate nell'analisi.

Tavola 20 – Follow-up 2005 - Abbinabili e non abbinabili per sesso e classe di età (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Maschi								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%		%	IC95%		%	IC95%	
0-14	26,0	23,1	28,9	14,8	14,5	15,1	15,0	14,7	15,4
15-24	10,5	8,4	12,4	11,0	10,8	11,3	11,0	10,7	11,3
25-34	17,1	14,6	19,7	15,1	14,8	15,4	15,1	14,8	15,5
35-44	15,6	13,2	18,0	16,8	16,5	17,1	16,8	16,4	17,1
45-54	9,9	8,0	12,0	13,4	13,1	13,7	13,3	13,0	13,6
55-64	9,4	7,5	11,4	12,2	12,0	12,5	12,2	11,9	12,5
65-69	3,0	1,9	4,3	5,3	5,1	5,5	5,3	5,1	5,5
70-74	3,9	2,5	5,2	4,6	4,4	4,8	4,6	4,4	4,7
75-79	1,8	0,8	2,6	3,5	3,3	3,6	3,5	3,3	3,6
>=80	2,8	1,6	3,8	3,3	3,1	3,4	3,2	3,1	3,4
Totale	100,0			100,0			100,0		

Classe di età	Femmine								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%		%	IC95%		%	IC95%	
0-14	20,9	18,4	23,4	13,3	13,1	13,6	13,5	13,2	13,8
15-24	11,0	9,0	12,9	9,9	9,6	10,1	9,9	9,7	10,2
25-34	18,1	15,7	20,4	14,1	13,8	14,3	14,1	13,8	14,4
35-44	15,8	13,6	18,1	15,6	15,4	15,9	15,6	15,3	16,0
45-54	8,1	6,4	9,8	13,0	12,8	13,3	12,9	12,7	13,2
55-64	9,4	7,6	11,3	12,2	12,0	12,5	12,2	11,9	12,4
65-69	2,6	1,7	3,7	5,8	5,6	6,0	5,7	5,5	5,9
70-74	3,5	2,4	4,7	5,3	5,1	5,5	5,3	5,1	5,5
75-79	4,1	2,9	5,5	4,7	4,5	4,8	4,7	4,5	4,8
>=80	6,5	5,0	8,1	6,0	5,8	6,2	6,0	5,8	6,2
Totale	100,0			100,0			99,9		

Classe di età	Totale								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%		%	IC95%		%	IC95%	
0-14	23,3	21,4	25,2	14,1	13,9	14,3	14,2	14,0	14,5
15-24	10,8	9,4	12,2	10,4	10,3	10,6	10,4	10,3	10,6
25-34	17,6	15,9	19,3	14,6	14,4	14,8	14,6	14,4	14,8
35-44	15,7	14,0	17,3	16,2	16,0	16,4	16,2	16,0	16,4
45-54	9,0	7,7	10,3	13,2	13,0	13,4	13,1	12,9	13,3
55-64	9,4	8,0	10,7	12,2	12,0	12,4	12,2	12,0	12,4
65-69	2,8	2,0	3,6	5,6	5,4	5,7	5,5	5,4	5,6
70-74	3,7	2,8	4,6	5,0	4,8	5,1	4,9	4,8	5,1
75-79	3,0	2,3	3,9	4,1	4,0	4,2	4,1	4,0	4,2
>=80	4,8	3,8	5,7	4,7	4,6	4,8	4,7	4,6	4,8
Totale	100,1			100,1			100,0		

Va tenuto conto che, per questa edizione dell'indagine, quasi la totalità del campione (98,3%) risulta "abbinabile" con l'archivio di mortalità e con quello delle dimissioni ospedaliere. Si ha dunque una affidabilità molto alta delle stime, ma una esiguità dei casi per i "non abbinabili", soprattutto quando si analizzano variabili con bassa prevalenza disaggregando per classi di età poco ampie. Ciò comporta una maggiore variabilità delle stime dei non abbinabili e quindi una maggiore ampiezza degli intervalli di confidenza.

Nonostante questo gli scostamenti dei "non abbinabili" dagli abbinabili e dall'intero campione, quando presenti, sono piuttosto contenute. Di seguito si riporta la distribuzione per genere e classe di età (Tavola 20), per titolo di studio ed età (Tavola 21), per territorio ed età (Tavola 22). I risultati dell'analisi sulle altre variabili considerate sono riportati nell'Appendice A.

Tavola 21 – Follow-up 2005 - Abbinabili e non abbinabili per classe di età e titolo di studio (valori percentuali e intervalli di confidenza al 95%)

Titolo di studio	Classe di età								
	25<=64								
	Non abbinabili		Abbinabili		Totale				
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	
Laurea, dottorato	19,3	16,9	21,8	12,6	12,4	12,9	12,7	12,5	13,0
Diploma di scuola media superiore	42,6	39,6	45,8	35,7	35,3	36,1	35,8	35,5	36,2
Licenza media	27,7	24,9	30,5	34,4	34,1	34,8	34,3	33,9	34,7
Licenza elementare	10,3	8,3	12,2	17,3	17,0	17,5	17,1	16,8	17,4
Totale	100,0			100,0			100,0		

Titolo di studio	Classe di età								
	65 anni e più								
	Non abbinabili		Abbinabili		Totale				
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	
Laurea, dottorato	3,6	1,4	6,2	4,0	3,7	4,3	4,0	3,7	4,3
Diploma di scuola media superiore	9,3	5,8	13,1	10,2	9,8	10,6	10,2	9,8	10,6
Licenza media	16,9	12,4	21,6	17,2	16,7	17,7	17,2	16,7	17,7
Licenza elementare	70,2	64,9	75,9	68,6	67,9	69,2	68,6	68,0	69,2
Totale	100,0			100,0			100,0		

Titolo di studio	Totale								
	Non abbinabili		Abbinabili		Totale				
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	
Laurea, dottorato	15,9	13,9	18,0	10,4	10,2	10,6	10,5	10,3	10,7
Diploma di scuola media superiore	35,4	32,8	38,1	29,2	28,9	29,5	29,3	29,0	29,6
Licenza media	25,4	23,0	27,8	30,0	29,7	30,3	29,9	29,6	30,2
Licenza elementare	23,3	21,0	25,7	30,4	30,1	30,7	30,2	29,9	30,5
Totale	100,0			100,0			100,0		

Tavola 22 – Follow-up 2005 - Abbinabili e non abbinabili per classe di età e ripartizione geografica (valori percentuali e intervalli di confidenza al 95%)

Ripartizione geografica	Classe di età								
	<=64								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%
Nord-ovest	24,2	21,9	26,4	26,0	25,7	26,3	26,0	25,7	26,3
Nord-est	16,7	15,1	18,3	18,6	18,4	18,9	18,6	18,4	18,8
Centro	28,6	26,4	30,7	18,6	18,4	18,9	18,8	18,6	19,1
Sud	19,4	17,8	21,1	25,0	24,7	25,3	24,9	24,6	25,1
Isole	11,1	9,5	12,5	11,7	11,5	11,9	11,7	11,5	11,9
Totale	100,0			100,0			100,0		

Ripartizione geografica	Classe di età								
	65 anni e più								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%
Nord-ovest	17,8	12,6	22,6	28,1	27,4	28,8	28,0	27,3	28,7
Nord-est	17,3	13,6	21,6	19,7	19,3	20,2	19,7	19,2	20,2
Centro	31,5	25,9	37,0	20,6	20,1	21,2	20,8	20,3	21,3
Sud	18,2	14,3	22,2	21,2	20,7	21,7	21,2	20,7	21,7
Isole	15,2	11,0	19,2	10,3	9,9	10,7	10,4	9,9	10,8
Totale	100,0			100,0			100,0		

Ripartizione geografica	Totale								
	Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%
Nord-ovest	23,3	21,3	25,4	26,4	26,1	26,7	26,4	26,0	26,7
Nord-est	16,8	15,4	18,3	18,9	18,7	19,1	18,8	18,6	19,0
Centro	29,0	27,0	31,0	19,0	18,8	19,3	19,2	19,0	19,5
Sud	19,2	17,7	20,8	24,3	24,0	24,5	24,2	23,9	24,4
Isole	11,7	10,3	13,1	11,4	11,2	11,6	11,4	11,2	11,6
Totale	100,0			100,0			100,0		

5.3 Controllo della *performance* del linkage con i decessi

Per valutare la qualità del *record linkage* con l'archivio delle cause di morte, è stato utilizzato il rapporto tra il numero di decessi osservati nel campione con il numero atteso, calcolato per ciascun anno di follow-up e per tutto il periodo. Sono stati esclusi dal calcolo gli anni in cui sono state effettuate le interviste (1999 e 2000 per la prima edizione dell'indagine e 2004-2005 per la seconda) in quanto, per questi anni, non è sempre stato possibile determinare con precisione il tempo di follow-up di ciascun individuo.

Il numero di decessi attesi per ogni anno è stato ottenuto sommando, per ciascun anno di *follow-up*, il prodotto tra i tassi di mortalità sesso ed età-specifici della popolazione italiana residente e il

numero effettivo di individui della coorte a rischio di morte per ogni anno, per sesso ed età (anni-persona)¹⁰.

I tassi di mortalità della popolazione italiana sono stati tratti dal database Health for All Italia [https://www.istat.it/it/archivio/14562], escludendo i residenti nelle province autonome di Trento e Bolzano quando per questi non erano disponibili le informazioni anagrafiche per la procedura di linkage (fino al 2003). Per il calcolo, sono state prese in considerazione classi di età decennali a partire dalla classe 0-14 anni, con l'ultima classe aperta di 75 e più anni.

I decessi attesi nell'anno t sono calcolati come la somma dei decessi attesi per sesso e ciascuna classe di età

$${}^tD_e = \sum_s \sum_x {}_{sx}{}^tD_e$$

dove

s è il sesso

x è la classe di età

I decessi attesi per ciascuna classe di età e sesso sono dati da:

$${}_{sx}{}^tD_e = {}_{sx}{}^tT_e {}_{sx}{}^tP_e$$

dove ${}_{sx}{}^tT_e$ è il tasso sesso ed età-specifico per la popolazione Italiana e ${}_{sx}{}^tP_e$ il numero effettivo di individui della coorte a rischio di morte per ogni anno, per sesso ed età (anni-persona). Per il calcolo degli anni-persona, per ogni anno t sono stati conteggiati tutti gli individui che non sono deceduti durante l'anno, mentre per i deceduti durante l'anno considerato è stata attribuita una sopravvivenza media di 6 mesi (0,5 anni).

Il rapporto osservati su attesi per ciascun anno t ${}^tR_{o/a}$ è dato da:

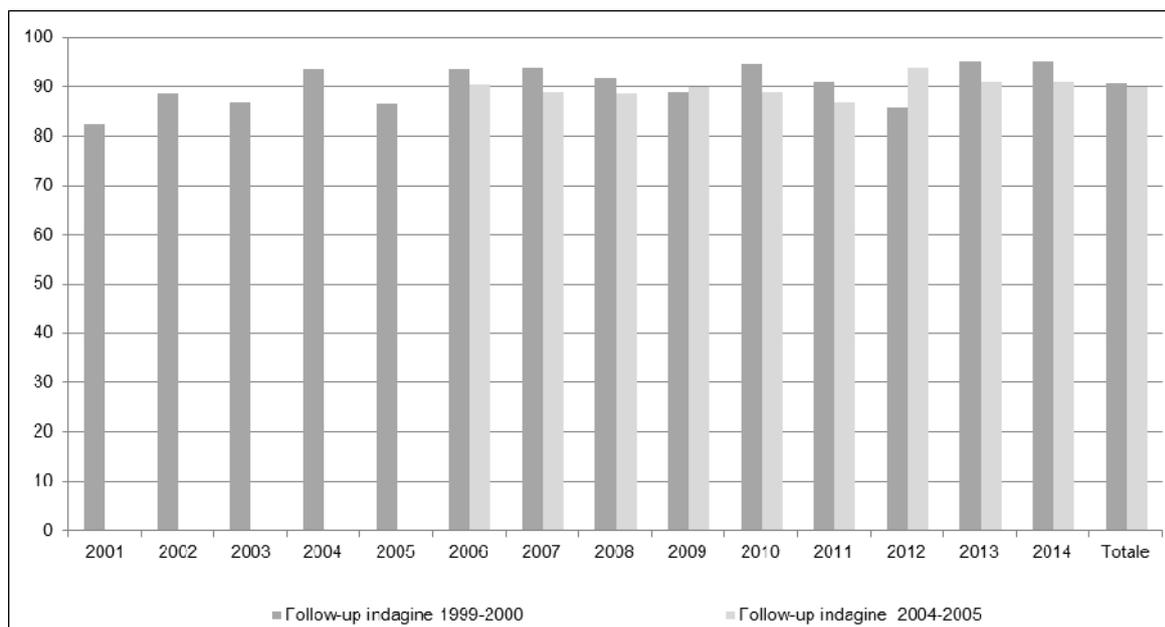
$${}^tR_{o/a} = \frac{{}^tD_o}{{}^tD_e} \quad \text{e il rapporto per l'intero periodo di follow-up è calcolato come } R_{o/a} = \frac{\sum_t {}^tD_o}{\sum_t {}^tD_e}$$

La Figura 4 mostra il rapporto tra il numero di morti osservati nelle due coorti (di tutte le età) e il numero atteso di decessi durante il follow-up. Il numero complessivo di decessi osservati nel periodo 2001-2014 per il Follow-up 2000 è stato di 18.330, di cui il 50,5% maschi; ciò corrisponde al 90% dei decessi attesi.

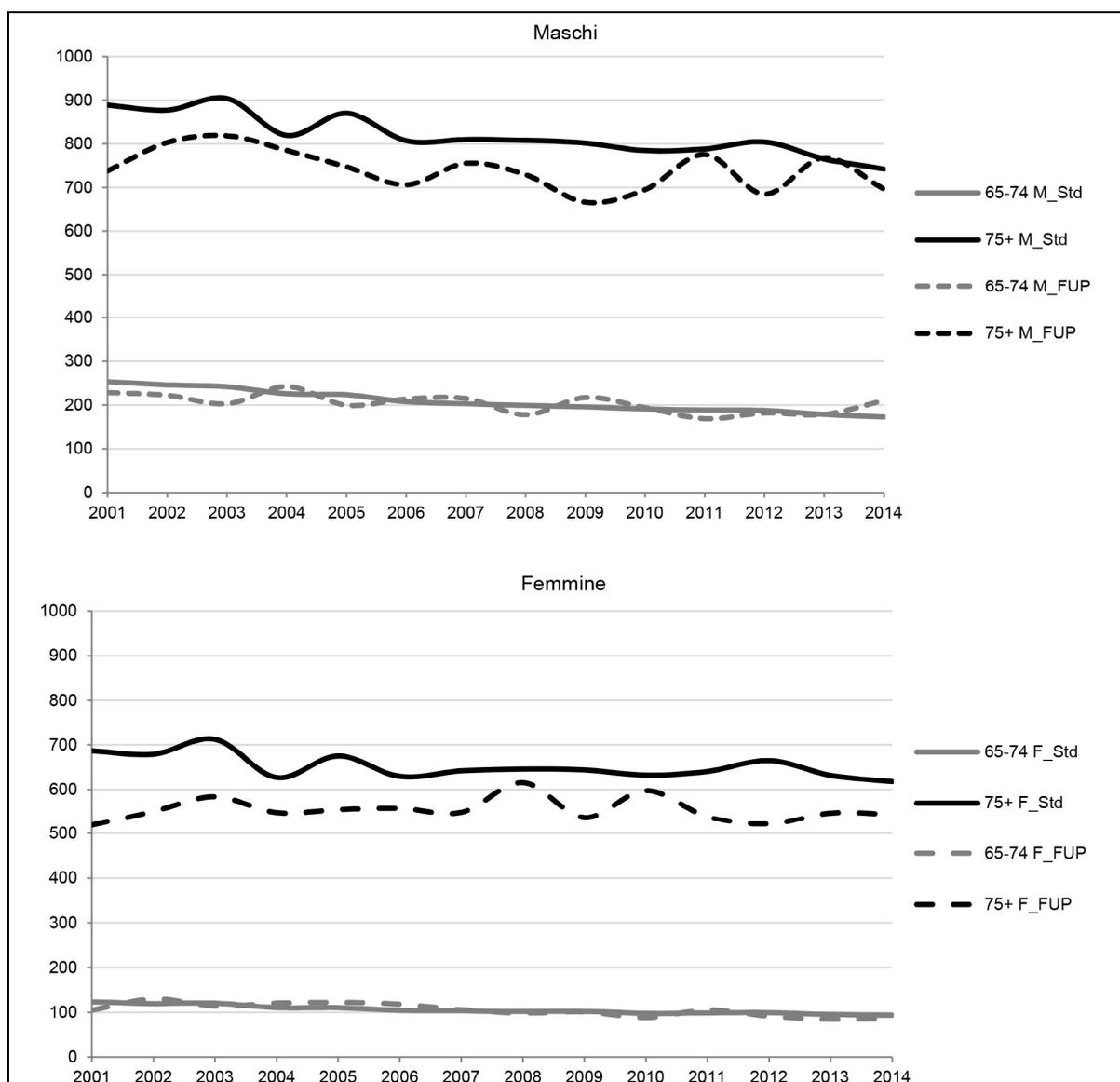
Il rapporto osservati-attesi ha avuto un andamento oscillante in tutto il periodo con tendenza alla crescita fra il 2001 e il 2006 raggiungendo valori del 94% con una successiva stabilizzazione. La crescita iniziale è verosimilmente dovuta al miglioramento del codice fiscale (chiave di *linkage*) registrato nell'archivio delle cause di morte a seguito del miglioramento generale della qualità di registrazione su supporto informatico delle informazioni alfanumeriche dei certificati di morte.

Un analogo valore del rapporto osservati-attesi si evidenzia per il periodo 2006-2014 per il Follow-up 2005 (in questo caso sono stati osservati nel periodo considerato 10.417 decessi) con andamenti meno variabili per anno rispetto a quanto rilevato per il Follow-up 2000.

¹⁰ I tassi di mortalità della popolazione italiana si riferiscono anche alle persone che vivono nelle istituzioni che non sono incluse, invece, tra gli intervistati dell'indagine sulla salute. I tassi di mortalità della popolazione generale potrebbero, quindi, differire da quelli del *follow-up* portando, in particolare, a sovrastimare il numero di decessi attesi specialmente nella popolazione anziana.

Figura 4 – Rapporto decessi osservati/attesi Follow-up 2000 e Follow-up 2005

Per gli anni 2003, 2005 e 2012, nel Follow-up 2000 si osservano rapporti osservati-attesi particolarmente bassi rispetto al valore medio (87% nel 2003, 86% per 2005 e 2012). È ipotizzabile che tale andamento sia dovuto a una differenza nella struttura per età del sottogruppo di individui effettivamente utilizzabili per il *record linkage* e quella del campione originale, rappresentativa della popolazione italiana che vive in famiglie. Infatti, nel caso del Follow-up 2000 il numero di record non abbinabili costituisce l'8% del totale del campione originale, ma con una percentuale diversa per classe di età e maggiore tra i più anziani: al di sopra di 75 anni sono stati esclusi dal linkage oltre il 10% dei soggetti intervistati (1.070 casi). La popolazione del Follow-up 2000 contiene quindi meno ultrasessantacinquenni rispetto al campione originario che ha una struttura per età rappresentativa della popolazione italiana in famiglia. Sebbene non statisticamente significativa, la minore presenza di ultrasessantacinquenni, gruppo con la mortalità più elevata, si traduce in un minor numero di decessi rispetto all'atteso. Tale fenomeno si accentua negli anni 2003, 2005 e 2012 in quanto in questi anni la mortalità generale, soprattutto nei più anziani, ha raggiunto dei picchi (Figure 5 e 6).

Figura 5 – Confronto dal 2001 al 2014 tra i tassi di mortalità della popolazione anziana e i tassi di mortalità degli anziani dell'indagine Salute 2000

Le oscillazioni descritte non si verificano per il Follow-up 2005 in quanto la quota di persone di 75 anni e più tra gli abbinabili è superiore o uguale a quella dei non abbinabili grazie al miglioramento della qualità delle informazioni dell'indagine sulla salute che ha consentito di raggiungere la quota di oltre il 98% degli abbinabili, quindi molto più elevata di quella ottenuta per il 2000 (92%).

Osservando il rapporto osservati-attesi per età e sesso, nel Follow-up 2000 si evidenziano valori più bassi tra le donne ultrasettantacinquenni per gli anni 2003, 2005 e 2012. Ciò non si riscontra nel Follow-up 2005 per l'anno 2012 (Appendice B).

5.4. Controllo della *performance* del linkage con i ricoveri

Per valutare la coerenza dei ricoveri osservati nel campione delle indagini sulla salute del 2000 e del 2005 con i tassi di ospedalizzazione osservati sulla popolazione negli anni di follow-up, è stato calcolato il rapporto tra ricoveri osservati e ricoveri attesi.

Innanzitutto, per entrambe le indagini, è stata conteggiata, per sesso e classi di età, la popolazione del campione anno per anno aggiornando l'età degli intervistati e decurtando per ciascun anno i deceduti. Un'analoga distribuzione è stata predisposta per i ricoveri osservati nei due campioni a seguito del *record linkage* con le SDO degli anni corrispondenti. Sono stati quindi calcolati i tassi di ospedalizzazione per classi di età e sesso del campione e confrontati con quelli della popolazione di riferimento.

Per effettuare l'analisi si è proceduto per ciascuna indagine al calcolo della popolazione residente in Italia corrispondente alla struttura per età che i due campioni assumevano negli anni successivi alla rilevazione (distribuzione per sesso e classi di età).

Per quanto riguarda l'indagine sulla salute del 2000, si è quindi considerata la popolazione dal 2001 (primo anno in cui erano disponibili i dati delle schede di dimissione ospedaliera) fino al 2014 per sesso e classe di età. Da essa, per ciascun anno, sono stati decurtati i minori nati successivamente al periodo di rilevazione dell'indagine (settembre 1999 – giugno 2000).

Per il calcolo della popolazione residente in Italia, dal 2001 al 2011 è stata utilizzata la popolazione al 1° gennaio di ciascun anno ricalcolata tra i due censimenti del 2001 e del 2011 (popolazione intercensuaria). Mentre dal 2012 al 2014 si è tenuto conto della popolazione residente al 1 gennaio. Per i dati di popolazione si è fatto riferimento alla banca dati Istat Demo (<http://demo.istat.it/>).

Per ottenere i tassi di ospedalizzazione della popolazione residente in Italia per ciascun anno, è stata calcolata, per sesso e per ogni classe di età, la popolazione media sommando la popolazione di ciascun anno a quella del successivo e dividendo per due. Anche la distribuzione per sesso e classe di età dei ricoveri per ciascun anno è stata decurtata dei minori nati dopo la rilevazione. Per i ricoveri della popolazione di riferimento si è fatto ricorso alla banca dati Armida delle SDO dal 2001 al 2014.

E' stato quindi calcolato, per ciascun anno, il tasso di ospedalizzazione età e genere specifico riportando i ricoveri alla popolazione media di riferimento di ciascun anno.

Si è dunque proceduto al calcolo dei ricoveri attesi applicando alla popolazione del campione il tasso di ospedalizzazione della popolazione media di riferimento di ciascun anno. Il calcolo è stato fatto per genere e classi di età e l'ammontare dei ricoveri attesi è la somma, per ciascun anno, dei ricoveri attesi per singola classe di età.

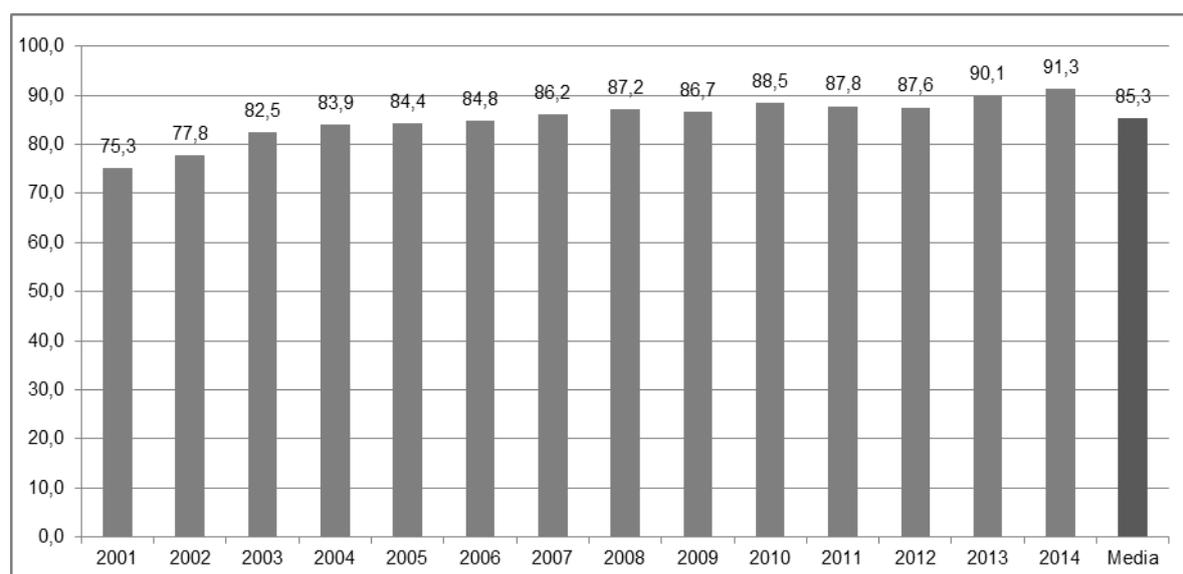
Un analogo procedimento è stato effettuato per l'indagine sulla salute 2004-2005, conteggiando la popolazione media di riferimento e i ricoveri dal 2004 al 2014.

Solo per il database costruito a partire dall'indagine sulla salute del 2000, è stata esclusa per il 2001 e per il 2002 la popolazione della Sardegna in quanto non erano disponibili le SDO per questi due anni¹¹.

Al termine della procedura per entrambi i Follow-up 2000 e 2005, sono stati calcolati i rapporti osservati/attesi (SHR) età e genere specifici per ciascun anno di follow-up (Appendice C).

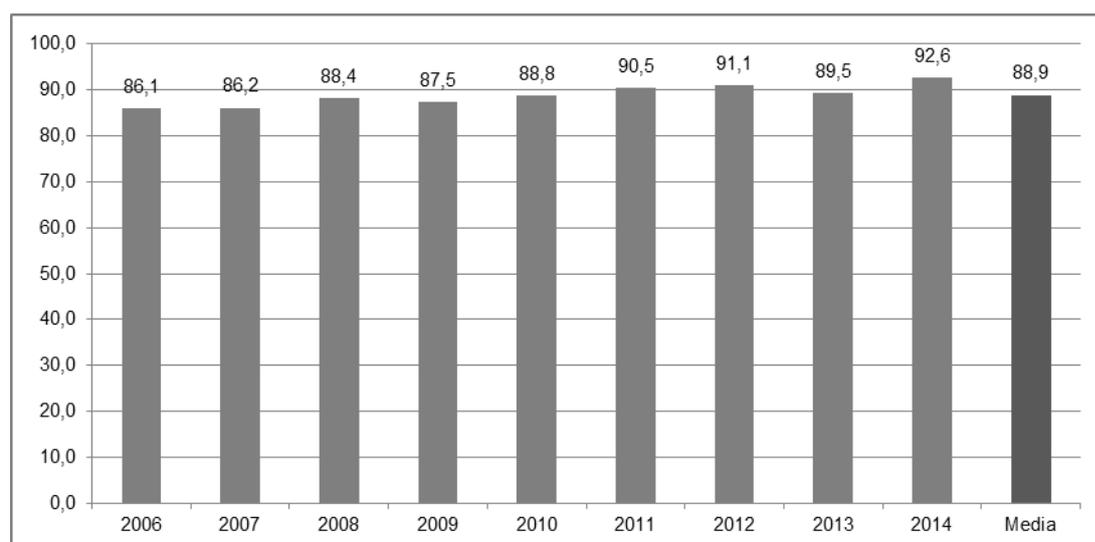
Per quanto riguarda il database costruito a partire dall'indagine sulla Salute del 2000, è possibile osservare il miglioramento del rapporto tra ricoveri osservati e attesi con il passare degli anni certamente imputabile al miglioramento della qualità del CF nelle schede di dimissione ospedaliera che influenza positivamente l'inclusione di tutti i ricoveri del campione e conseguentemente la riduzione della differenza tra tassi di ospedalizzazione della popolazione di riferimento e quelli del campione (Figura 6). Dal 2013 il rapporto supera il 90%.

¹¹ Il Ministero della Salute non li ha forniti perché per questi due anni non era presente il CF.

Figura 6 – Rapporto ricoveri osservati/attesi - Follow-up 2000

Per quanto riguarda il rapporto osservati/attesi dei ricoveri abbinati all'indagine sulla Salute del 2005 il valore medio raggiunge quasi l'89%, quasi 4 punti percentuali in più rispetto al Follow-up 2000 (Figura 7). Questo avviene in parte a causa della peggiore *performance* nei primi anni di linkage del Follow-up 2000 ma anche presumibilmente per il miglioramento della qualità del CF nell'indagine sulla Salute del 2005 grazie al nuovo sistema di rilevazione attivato in Istat.

Il calcolo dei ricoveri attesi e quindi del rapporto con quelli osservati è stato effettuato a partire dal 2006. Ciò in quanto la rilevazione è stata effettuata in quattro trimestri successivi a partire da dicembre 2004 fino a settembre 2005 e ciò non consentiva di determinare con precisione il tempo di follow-up di ciascun individuo.

Figura 7 – Rapporto ricoveri osservati/attesi - Follow-up 2005

5.5. Confronto tra proporzione di deceduti e ricoverati nei follow-up delle indagini 1999-2000 e 2004-2005.

Nella Tavola 23 sono riportate le numerosità non pesate del campione abbinabile dell'indagine sulla salute del 1999-2000 e di quella del 2004-2005, la proporzione di deceduti complessivi e quella di prima ospedalizzazione per patologia dei soggetti ricoverati osservati nei primi dieci anni di follow-up delle indagini Istat sulla salute 1999-2000 e 2004-2005. Si evidenzia come la proporzione di deceduti in entrambi i generi sia molto simile nei due follow-up (follow-up 2000-2009: 9,2% negli uomini e 8.1% nelle donne; follow-up 2005-2014: 9.1% negli uomini e 8.5% nelle donne). Un analogo risultato si osserva per la proporzione di soggetti ricoverati per specifiche cause nei follow-up delle due indagini.

Ciò dimostra come le procedure di *record linkage* dei soggetti intervistati nelle due indagini con l'archivio nazionale di mortalità e con quello dei ricoveri ospedalieri abbiano funzionato in maniera sovrapponibile per entrambe le coorti delle due indagini sulla salute.

Tavola 23 – Confronto tra Follow-up 2000 e Follow-up 2005 per decessi e principali cause di ricovero - uomini (valori assoluti, percentuali e p-value relativo al test χ^2 per il confronto tra le frequenze tra le due indagini)

	Follow-up 2000 ^a	% casi	Follow-up 2005 ^b	% casi	Test χ^2 p-value ^c
Popolazione	62895		60917		
Decessi	5807	9.2%	5541	9.1%	0.404
Ricoveri per:					
Malattia cardiaca coronarica	2215	0,035	2421	0,04	<0.001
Altre malattie cardiache	432	0.7%	424	0.7%	0.846
Malattie cerebrovascolari	2.004	3.2%	2.095	3.4%	0.013
Demenza	147	0.2%	139	0.2%	0.839
Asma e broncopneumopatia cronica ostruttiva	181	0.3%	156	0.3%	0.285
Epatite	109	0.2%	42	0.1%	<0.001
Depressione	88	0.1%	77	0.1%	0.515
Cirrosi	252	0.4%	215	0.4%	0.171
Psicosi	252	0.4%	271	0.4%	0.231
Emicrania	56	0.1%	35	0.1%	0.040
Diabete	543	0.9%	451	0.7%	0.015
Colite cronica	122	0.2%	140	0.2%	0.170
Artrosi	686	1.1%	814	1.3%	<0.001
Artrite reumatoide	51	0.1%	57	0.1%	0.457
Fratture del femore	411	0.7%	466	0.8%	0.019
Fratture cranio-vertebrali	60	0.1%	55	0.1%	0.768
Ipertensione maligna	446	0.7%	414	0.7%	0.532
Alcolismo	118	0.2%	95	0.2%	0.179
Disordini neuro e mio degenerativi (SLA, sclerosi multipla, altre mielopatie, miastenia grave, distrofia muscolare)	116	0.2%	98	0.2%	0.318
Disordini del tessuto connettivo	18	0.0%	31	0.1%	0.049
Tumori maligni	3.709	5.9%	3.972	6.5%	<0.001
Malattia di Parkinson	90	0.1%	87	0.1%	0.990
Epilessia	209	0.3%	205	0.3%	0.898

^a Follow-up indagine salute 1999-2000 - Periodo di osservazione 2000-2009

^b Follow-up indagine salute 2004-2005 - Periodo di osservazione 2005-2014

^c test χ^2 per il confronto tra proporzioni di decessi e delle singole cause di ricovero tra le due indagini

Tavola 23 segue – Confronto tra Follow-up 2000 e Follow-up 2005 per decessi e principali cause di ricovero - donne (valori assoluti, percentuali e p-value relativo al test χ^2 per il confronto tra le frequenze tra le due indagini)

	Follow-up 2000 ^a	% casi	Follow-up 2005 ^b	% casi	Test χ^2 p-value ^c
Popolazione	65.923		64.933		
Decessi	5.349	8.1%	5.509	8.5%	0.015
Ricoveri per:					
Malattia cardiaca coronarica	1.172	1,80%	1.228	1,90%	0.127
Altre malattie cardiache	301	0.5%	297	0.5%	0.983
Malattie cerebrovascolari	1.914	2.9%	2.004	3.1%	0.052
Demenza	214	0.3%	223	0.3%	0.555
Asma e broncopneumopatia cronica ostruttiva	196	0.3%	160	0.2%	0.077
Epatite	83	0.1%	53	0.1%	0.013
Depressione	219	0.3%	168	0.3%	0.014
Cirrosi	168	0.3%	147	0.2%	0.294
Psicosi	289	0.4%	327	0.5%	0.085
Emicrania	140	0.2%	119	0.2%	0.236
Diabete	467	0.7%	376	0.6%	0.003
Colite cronica	124	0.2%	113	0.2%	0.549
Artrosi	1.185	1.8%	1.319	2.0%	0.002
Artrite reumatoide	139	0.2%	108	0.2%	0.064
Fratture del femore	974	1.5%	1.177	1.8%	<0.001
Fratture cranio-vertebrali	20	0.0%	37	0.1%	0.021
Ipertensione maligna	515	0.8%	508	0.8%	0.981
Alcolismo	34	0.1%	37	0.1%	0.675
Disordini neuro e mio degenerativi (SLA, sclerosi multipla, altre mielopatie, miastenia grave, distrofia muscolare)	161	0.2%	139	0.2%	0.254
Disordini del tessuto connettivo	135	0.2%	128	0.2%	0.757
Tumori maligni	3.116	4.7%	3.429	5.3%	<0.001
Malattia di Parkinson	77	0.1%	64	0.1%	0.315
Epilessia	204	0.3%	176	0.3%	0.197

^a Follow-up indagine salute 1999-2000 - Periodo di osservazione 2000-2009

^b Follow-up indagine salute 2004-2005 - Periodo di osservazione 2005-2014

^c test χ^2 per il confronto tra proporzioni di decessi e delle singole cause di ricovero tra le due indagini

6. Considerazioni finali

L'esperienza di integrazione delle fonti condotta ha dato buoni risultati. La ricerca scientifica ha incominciato a valorizzare questa nuova fonte informativa con risultati promettenti (Marinacci *et al.*, 2011; Marinacci *et al.*, 2013; Federico *et al.*, 2013; D'Errico *et al.*, 2016; Piccinelli *et al.*, 2016; Piccinelli *et al.*, 2018)¹².

Parallelamente agli obiettivi scientifici sono molte le potenzialità di uso dello studio longitudi-

¹² Per ragioni di tutela dei dati personali l'utilizzo di questa fonte informativa è riservato esclusivamente ai soggetti firmatari della convenzione che hanno titolarità per il trattamento dei dati. I lavori fin qui svolti sono stati infatti realizzati da Ministero della Salute, Servizio di Epidemiologia ASL TO3 del Piemonte e Istat. Questo aspetto al momento costituisce una criticità per la possibilità di accesso e quindi di analisi da parte di altri attori interessati.

nale italiano per la programmazione e la valutazione delle politiche in campo preventivo e sanitario, di cui si citano di seguito alcuni esempi di particolare attualità.

Il welfare aziendale è in fase di particolare espansione, ma sono poco note le implicazioni per la salute delle soluzioni su cui si applica. Ad esempio, per quanto riguarda i problemi di conciliazione tra famiglia e lavoro, lo studio longitudinale italiano, consentendo di osservare in modo prospettico gli esiti di salute (attraverso l'integrazione con i dati di ospedalizzazione) offre l'opportunità inedita di fornire informazioni sui rischi per la morbosità delle lavoratrici con figli per valutare i bisogni e l'impatto degli interventi di conciliazione. Il nuovo Piano Nazionale Cronicità (PNC) sollecita le regioni e le aziende sanitarie a sviluppare formule di stratificazione del rischio che permettano una migliore e più personalizzata presa in carico dell'assistenza primaria e specialistica: la ricchezza di informazioni su salute, condizione socio-economica, condizione familiare, morbosità e mortalità dello studio longitudinale italiano permette di fornire nuovi strumenti per la stratificazione del rischio a livello di comunità e individuale. Lo stesso PNC affronta sfide di programmazione sull'assistenza della cronicità in una popolazione che invecchia in assenza di chiare prove circa la fondatezza delle ipotesi contrastanti sulla compressione o espansione della morbosità che sarebbe in corso negli ultimi anni di vita. Su questo tema è fondamentale far luce per prevedere l'andamento del fabbisogno di assistenza e di personale e dei relativi costi per i prossimi decenni; ancora una volta lo studio longitudinale italiano può fornire un contributo inedito in questa direzione. Nell'agenda della programmazione sanitaria delle regioni è all'ordine del giorno la necessità di una revisione dei criteri di riparto del fondo sanitario che sia sensibile alle variazioni regionali nei principali determinanti del fabbisogno; proprio su questi determinanti lo studio longitudinale offre possibilità di analisi innovative (morbosità, invecchiamento e deprivazione). Infine allo scadere del Piano Nazionale di Prevenzione 2014-2018 sono iniziati i lavori per l'elaborazione del nuovo piano e lo studio longitudinale italiano potrebbe fornire nuovi input conoscitivi sull'impatto atteso sui Disability Adjusted Life Years (DALYs) e sui costi sanitari dei diversi determinanti considerati dal piano, al fine di fare delle valutazioni di priorità più circostanziate e giustificate dai dati.

Per tutte queste ragioni lo studio longitudinale italiano è un patrimonio informativo che andrà aggiornato ed arricchito con nuovi sviluppi. La metodologia e le procedure sperimentate saranno estese anche all'indagine Istat sulla salute condotta nel 2012-2013 per la quale, come per le precedenti edizioni, il campione è stato ampliato a 60.000 famiglie. Sarà inoltre condotta una sperimentazione per integrare con i dati di mortalità e ospedalizzazione l'indagine "Condizione e Integrazione sociale dei cittadini stranieri" realizzata nel 2011-2012 nell'ambito della quale sono state raccolte informazioni sulle condizioni di salute, sui fattori di rischio, sul ricorso alle cure e sull'accessibilità dei servizi sanitari per gli immigrati. Per entrambe le indagini è stato già avviato il lavoro per il *record linkage*. E' anche prevista l'integrazione dei dati relativi a decessi e ricoveri con l'indagine Istat "Aspetti della vita quotidiana (AVQ)", che permetterà di sottomettere a Follow-up prospettico coorti di intervistati per i quali sono state rilevate informazioni su fattori di rischio per la salute che non sono disponibili nelle indagini sulla salute, come l'alimentazione e l'uso di alcool. È inoltre in programma l'analisi di fattibilità per l'integrazione dei due Follow-up con le informazioni relative alle prestazioni farmaceutiche e alle prestazioni di assistenza specialistica ambulatoriale (prestazioni terapeutiche e riabilitative, diagnostica strumentale e di laboratorio) con una sperimentazione con alcune regioni (Piemonte, Lazio, Liguria, Emilia-Romagna, Sicilia e Puglia). Ciò consentirebbe di studiare l'influenza della condizione socio-economica sul consumo di servizi sanitari.

7. Appendici

Appendice A

Comparazione abbinabili, non abbinabili, campione - Follow-up 2000 e Follow-up 2005

Follow-up 2000

Follow-up 2000 - Abbinabili e non abbinabili per classe di età e risorse economiche (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Risorse economiche																	
	Ottime/adequate						Scarse/insufficienti											
	Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale							
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%						
<=24	72,0	70,0	74,1	70,8	70,1	71,5	70,9	70,2	71,6	28,0	26,0	30,0	29,2	28,5	29,9	29,1	28,5	29,8
25-44	73,2	71,1	75,2	74,2	73,6	74,8	74,1	73,6	74,7	26,8	24,8	28,8	25,8	25,2	26,4	25,9	25,3	26,4
45-64	71,9	69,5	74,2	73,4	72,8	74,1	73,3	72,7	74,0	28,1	25,8	30,5	26,6	25,9	27,2	26,7	26,0	27,3
>=65	66,9	64,2	69,6	68,0	67,1	68,8	67,9	67,0	68,7	33,1	30,4	35,8	32,0	31,2	32,9	32,1	31,3	33,0
Totale	71,4	70,2	72,5	72,0	71,7	72,4	72,0	71,3	72,6	28,6	27,5	29,8	28,0	27,6	28,3	28,0	27,6	28,4

Follow-up 2000 - Abbinabili e non abbinabili per classe di età e condizione lavorativa (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Condizione lavorativa																										
	Occupati			In cerca di occupazione			Ritirati dal lavoro/Inabili al lavoro/In altra condizione																				
	Non abbinabili		Abbinabili	Totale		Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale											
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%											
15-24	24,6	21,6	27,5	23,9	23,0	24,9	24,0	23,1	24,9	18,5	15,9	21,2	15,6	14,8	16,4	15,8	15,0	16,6	56,9	53,5	60,3	60,5	59,4	61,6	60,2	59,1	61,2
25-44	68,7	66,5	70,8	70,6	70,0	71,3	70,5	69,9	71,1	12,2	10,8	13,7	10,8	10,4	11,2	10,9	10,5	11,3	19,0	17,2	20,8	18,6	18,0	19,1	18,6	18,1	19,1
45-64	48,2	45,5	50,8	47,4	46,7	48,2	47,5	46,7	48,2	3,4	2,4	4,3	3,4	3,1	3,7	3,4	3,1	3,7	48,4	45,8	51,0	49,2	48,4	50,0	49,1	48,4	49,9
>=65	3,1	2,1	4,0	2,1	1,9	2,4	2,2	1,9	2,4	-	-	-	-	-	-	-	-	-	96,9	95,9	97,9	97,9	97,6	98,1	97,8	97,5	98,1
Totale	41,8	40,4	43,2	43,3	42,9	43,7	43,2	42,8	43,6	8,1	7,3	8,8	7,1	6,8	7,3	7,1	6,9	7,4	50,1	48,7	51,5	49,7	49,2	50,1	49,7	49,3	50,1

Follow-up 2000 - Abbinabili e non abbinabili per classe di età e salute percepita (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Salute percepita																										
	Male o molto male			Discretamente						Bene o molto bene																	
	Non abbinabili		Abbinabili	Totale		Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale											
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%											
0-14	0,9	0,3	1,3	0,8	0,6	1,0	0,8	0,7	1,0	6,5	5,1	7,9	7,4	6,9	8,0	7,3	6,9	7,8	92,6	91,1	94,2	91,8	91,2	92,3	91,8	91,3	92,4
15-24	1,6	0,7	2,3	1,2	1,0	1,4	1,2	1,0	1,5	10,8	8,6	12,8	12,2	11,5	13,0	12,1	11,4	12,8	87,6	85,3	89,9	86,6	85,8	87,3	86,7	85,9	87,4
25-34	1,4	0,7	2,0	1,8	1,5	2,0	1,8	1,5	2,0	19,1	16,7	21,5	19,0	18,3	19,8	19,0	18,3	19,8	79,6	77,0	82,0	79,2	78,4	80,0	79,2	78,4	80,0
35-44	3,7	2,5	4,9	2,6	2,3	2,9	2,7	2,4	3,0	27,8	24,8	30,7	28,5	27,6	29,4	28,4	27,6	29,3	68,5	65,4	71,6	68,9	68,0	69,8	68,9	68,0	69,7
45-54	5,2	3,8	6,8	5,5	5,1	6,0	5,5	5,1	6,0	42,0	38,5	45,4	41,6	40,5	42,6	41,6	40,6	42,6	52,8	49,4	56,4	52,9	51,8	53,9	52,9	51,9	53,9
55-64	11,1	8,7	13,4	10,2	9,6	10,9	10,3	9,7	10,9	50,1	46,2	53,8	52,1	50,9	53,2	51,9	50,8	53,0	38,8	35,0	42,4	37,7	36,6	38,8	37,8	36,7	38,8
65-74	17,7	14,8	20,8	17,3	16,4	18,2	17,4	16,5	18,2	58,3	54,3	62,2	59,0	57,9	60,2	59,0	57,8	60,1	24,0	20,6	27,3	23,6	22,6	24,6	23,6	22,7	24,6
>=75	34,1	30,4	38,0	30,2	28,9	31,5	30,6	29,3	31,8	85,4	82,4	88,4	85,4	82,4	88,4	85,4	82,4	88,4	57,1	53,5	60,1	57,1	53,5	60,1	57,1	53,5	60,1
Totale	7,6	6,9	8,2	6,9	6,7	7,1	7,0	6,8	7,2	29,6	28,5	30,8	31,8	31,5	32,2	31,6	31,2	32,1	62,8	61,5	64,0	61,3	60,9	61,7	61,4	60,8	62,0

Follow-up 2000 - Abbinabili e non abbinabili per classe di età, genere e disabilità (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Maschi																	
	Non disabile									Disabile								
	Non abbinabili			Abbinabili			Totale			Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%
6-34	98,8	98,2	99,4	99,0	98,8	99,2	99,0	98,8	99,2	1,2	0,6	1,8	1,0	0,8	1,1	1,0	0,8	1,2
35-64	98,9	98,4	99,6	98,3	98,0	98,5	98,3	98,1	98,5	1,1	0,5	1,7	1,7	1,5	2,0	1,7	1,5	1,9
65-69	95,0	92,1	98,7	93,6	92,4	94,8	93,7	92,6	94,8	5,0	1,5	7,7	6,4	5,3	7,5	6,3	5,2	7,3
70-74	88,4	82,8	94,4	90,4	88,8	91,9	90,2	88,7	91,6	11,6	5,8	17,0	9,6	8,2	11,1	9,8	8,3	11,2
75-79	80,4	73,2	87,8	86,1	84,0	88,2	85,6	83,6	87,6	19,6	12,3	26,6	13,9	11,9	15,9	14,4	12,4	16,3
>=80	60,3	50,7	69,8	61,4	58,1	64,8	61,3	58,1	64,5	39,7	30,2	49,2	38,6	35,3	41,9	38,7	35,5	41,8
Totale	96,3	95,5	97,0	96,6	96,4	96,8	96,6	96,4	96,8	3,7	3,0	4,4	3,4	3,2	3,6	3,4	3,2	3,6

Classe di età	Femmine																	
	Non disabile									Disabile								
	Non abbinabili			Abbinabili			Totale			Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%
6-34	98,8	98,1	99,4	98,9	98,7	99,1	98,9	98,7	99,1	1,2	0,7	1,9	1,1	0,9	1,3	1,1	0,9	1,3
35-64	97,0	96,0	98,0	97,9	97,7	98,2	97,9	97,6	98,1	3,0	2,1	3,9	2,1	1,9	2,3	2,1	1,9	2,4
65-69	91,6	87,3	95,2	92,5	91,5	93,7	92,5	91,4	93,6	8,4	4,4	11,7	7,5	6,4	8,6	7,5	6,4	8,6
70-74	88,2	83,3	93,0	86,7	85,1	88,3	86,8	85,3	88,3	11,8	7,1	16,6	13,3	11,7	14,8	13,2	11,7	14,6
75-79	76,0	70,6	82,2	77,1	75,1	79,2	77,0	75,1	79,0	24,0	18,5	30,1	22,9	20,8	24,9	23,0	21,1	24,9
>=80	46,8	40,1	53,7	48,1	45,7	50,5	48,0	45,7	50,3	53,2	46,3	59,9	51,9	49,4	54,3	52,0	49,7	54,3
Totale	92,4	91,5	93,4	93,9	93,7	94,2	93,8	93,5	94,1	7,6	6,6	8,5	6,1	5,8	6,3	6,2	5,9	6,4

Classe di età	Totale																	
	Non disabile									Disabile								
	Non abbinabili			Abbinabili			Totale			Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%
6-34	98,8	98,3	99,2	99,0	98,8	99,1	99,0	98,8	99,1	1,2	0,8	1,7	1,0	0,9	1,2	1,0	0,9	1,2
35-64	98,0	97,4	98,6	98,1	97,9	98,3	98,1	97,9	98,2	2,0	1,5	2,6	1,9	1,8	2,1	1,9	1,8	2,1
65-69	93,1	90,4	95,8	93,0	92,2	93,8	93,0	92,2	93,8	6,9	4,3	9,5	7,0	6,2	7,8	7,0	6,2	7,7
70-74	88,3	84,5	92,1	88,3	87,2	89,5	88,3	87,3	89,4	11,7	8,0	15,4	11,7	10,6	12,7	11,7	10,6	12,7
75-79	77,5	72,7	82,0	80,7	79,2	82,2	80,4	79,0	81,9	22,5	18,0	27,2	19,3	17,8	20,7	19,6	18,1	21,0
>=80	51,1	45,3	56,6	52,5	50,5	54,5	52,3	50,4	54,2	48,9	43,4	54,7	47,5	45,5	49,5	47,7	45,8	49,6
Totale	94,3	93,6	94,9	95,2	95,1	95,4	95,2	95,0	95,3	5,7	5,2	6,4	4,8	4,6	4,9	4,8	4,7	5,0

Follow-up 2000 - Abbinabili e non abbinabili per classe di età e abitudine al fumo (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Abitudine al fumo																										
	Sì									No, ma fumava in passato									No, non ha mai fumato								
	Non abbinabili			Abbinabili			Totale			Non abbinabili			Abbinabili			Totale			Non abbinabili			Abbinabili			Totale		
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%			
14-24	22,2	19,5	24,9	22,2	21,3	23,0	22,2	21,3	23,0	4,7	3,4	6,1	5,4	4,9	5,8	5,3	4,9	5,7	73,1	70,3	76,2	72,5	71,5	73,4	72,5	71,6	73,5
25-44	30,5	28,4	32,6	33,1	32,5	33,8	32,9	32,3	33,6	14,7	13,1	16,3	15,6	15,1	16,2	15,6	15,1	16,1	54,8	52,5	57,1	51,2	50,5	51,9	51,5	50,8	52,2
45-64	27,2	24,9	29,5	26,4	25,8	27,1	26,5	25,8	27,2	21,6	19,5	23,8	24,6	24,0	25,3	24,4	23,8	25,0	51,2	48,6	53,8	48,9	48,1	49,7	49,1	48,3	49,8
>=65	9,2	7,6	10,9	10,9	10,4	11,5	10,8	10,2	11,3	24,5	22,0	26,9	28,4	27,6	29,2	28,0	27,3	28,8	66,3	63,6	69,0	60,7	59,8	61,6	61,2	60,3	62,0
Italia	23,5	22,4	24,7	25,1	24,7	25,4	24,9	24,6	25,3	17,1	16,1	18,1	19,3	19,0	19,7	19,1	18,8	19,5	59,4	58,0	60,8	55,6	55,2	56,1	55,9	55,5	56,3

Follow-up 2000 - Abbinabili e non abbinabili per classe di età e Body Mass Index (valori percentuali e intervalli di confidenza al 95%)

Classe di età	BMI in classi																	
	Sottopeso						Normopeso											
	Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale							
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%						
18-24	11,5	9,1	14,1	10,5	9,7	11,3	10,6	9,9	11,4	75,8	72,3	79,2	75,3	74,1	76,4	75,3	74,2	76,4
25-34	5,6	4,1	6,9	5,6	5,2	6,0	5,6	5,2	6,0	70,0	67,1	72,8	69,5	68,6	70,4	69,5	68,7	70,4
35-44	2,3	1,4	3,3	2,4	2,1	2,7	2,4	2,1	2,7	60,3	57,0	63,5	59,0	58,0	60,0	59,1	58,2	60,0
45-54	1,6	0,8	2,4	1,4	1,1	1,6	1,4	1,2	1,6	48,2	44,6	51,7	47,4	46,4	48,5	47,5	46,5	48,5
55-64	1,4	0,5	2,1	1,3	1,1	1,5	1,3	1,1	1,5	43,7	39,9	47,5	41,4	40,3	42,5	41,5	40,5	42,6
65-74	1,9	1,0	3,1	1,4	1,1	1,7	1,4	1,2	1,7	41,9	37,9	45,7	41,3	40,1	42,5	41,4	40,2	42,5
>=75	5,4	3,8	7,3	4,4	3,8	4,9	4,5	3,9	5,0	48,1	44,0	52,0	47,9	46,5	49,3	47,9	46,6	49,3
Totale	4,1	3,6	4,7	3,6	3,4	3,7	3,6	3,5	3,8	56,5	55,1	57,9	55,0	54,6	55,5	55,1	54,7	55,6

Classe di età	BMI in classi																	
	Sovrappeso						Obeso											
	Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale							
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%						
18-24	11,5	9,1	14,1	12,6	11,7	13,4	12,5	11,7	13,3	1,2	0,3	1,9	1,6	1,3	2,0	1,6	1,3	1,9
25-34	20,7	18,2	23,1	21,4	20,6	22,2	21,3	20,5	22,1	3,7	2,6	4,9	3,5	3,2	3,9	3,6	3,2	3,9
35-44	29,8	26,8	32,8	31,6	30,7	32,5	31,4	30,6	32,3	7,6	5,9	9,3	7,0	6,5	7,5	7,1	6,6	7,5
45-54	39,0	35,6	42,4	39,0	37,9	40,0	39,0	38,0	40,0	11,2	9,1	13,5	12,2	11,6	12,9	12,2	11,5	12,8
55-64	40,1	36,2	43,7	43,1	42,0	44,2	42,9	41,9	44,0	14,8	12,2	17,5	14,2	13,4	14,9	14,2	13,5	15,0
65-74	44,0	40,1	47,9	43,2	42,0	44,4	43,3	42,1	44,4	12,2	9,6	14,7	14,0	13,2	14,9	13,9	13,1	14,7
>=75	37,0	33,1	40,8	37,4	36,0	38,8	37,4	36,1	38,6	9,5	7,2	11,8	10,3	9,5	11,2	10,2	9,4	11,0
Totale	31,1	29,8	32,4	32,6	32,1	33,0	32,4	32,0	32,8	8,3	7,5	9,1	8,9	8,6	9,1	8,8	8,6	9,1

Follow-up 2005

Follow-up 2005 - Abbinabili e non abbinabili per classe di età e risorse economiche (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Risorse economiche																	
	Ottime/adequate						Scarse/insufficienti											
	Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale							
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%						
<=24	66,3	62,8	70,0	66,6	66,0	67,1	66,6	66,0	67,1	33,7	30,2	37,5	33,4	32,9	34,0	33,4	32,9	34,0
25-44	71,9	68,4	75,4	70,0	69,6	70,5	70,1	69,6	70,5	28,1	24,6	31,6	30,0	29,5	30,5	29,9	29,5	30,4
45-64	77,2	73,0	81,9	71,3	70,8	71,8	71,4	70,8	71,9	22,8	18,1	27,0	28,7	28,2	29,3	28,6	28,1	29,2
>=65	57,7	51,9	63,8	63,4	62,8	64,1	63,4	62,7	64,0	42,3	36,1	48,1	36,6	35,9	37,2	36,6	36,0	37,3
Totale	68,9	66,9	71,0	68,2	68,0	68,5	68,2	67,8	68,7	31,1	29,0	33,1	31,8	31,5	32,0	31,8	31,5	32,1

Follow-up 2005 - Abbinabili e non abbinabili per classe di età e condizione lavorativa (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Condizione lavorativa																										
	Occupati				In cerca di occupazione				Ritirati dal lavoro/Inabili al lavoro/In altra condizione																		
	Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale										
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%									
15-24	29,9	23,5	36,5	25,6	24,8	26,4	25,7	24,9	26,5	18,7	12,8	23,8	13,6	13,0	14,3	13,7	13,1	14,3	51,4	43,8	57,8	60,8	59,9	61,7	60,6	59,7	61,5
25-44	75,3	71,8	78,5	74,5	74,1	75,0	74,6	74,1	75,0	10,6	8,1	13,0	8,5	8,2	8,8	8,5	8,2	8,8	14,1	11,3	16,8	17,0	16,6	17,4	16,9	16,5	17,3
45-64	56,9	51,6	62,1	54,1	53,6	54,7	54,2	53,6	54,7	3,2	1,4	5,5	2,7	2,5	2,9	2,7	2,5	2,9	39,9	34,5	44,9	43,1	42,6	43,7	43,1	42,5	43,7
>=65	3,6	1,4	6,2	3,5	3,3	3,8	3,5	3,3	3,8	-	-	-	-	-	-	-	-	-	96,3	93,9	98,5	96,4	96,1	96,6	96,4	96,2	96,6
Totale	51,1	48,6	53,6	46,6	46,3	46,9	46,7	46,4	47,0	8,0	6,6	9,4	5,5	5,4	5,7	5,6	5,4	5,7	40,9	38,3	43,3	47,9	47,6	48,2	47,7	47,4	48,0

Follow-up 2005 - Abbinabili e non abbinabili per classe di età e salute percepita (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Giudizio sulla salute percepita																										
	Bene o molto bene					Discretamente					Male o molto male																
	Non abbinabili		Abbinabili		Totale	Non abbinabili		Abbinabili		Totale	Non abbinabili		Abbinabili		Totale												
	%	IC95%	%	IC95%	%	%	IC95%	%	IC95%	%	%	IC95%	%	IC95%	%												
0-14	97,3	95,8	98,8	94,7	94,4	95,1	94,8	94,5	95,2	2,7	1,1	4,3	4,9	4,5	5,2	4,8	4,5	5,1	.	-	-	0,4	0,3	0,5	0,4	0,3	0,5
15-24	88,9	83,9	92,8	90,7	90,2	91,3	90,7	90,2	91,2	10,7	6,4	15,3	8,8	8,2	9,3	8,8	8,3	9,3	0,4	-0,5	2,2	0,5	0,4	0,7	0,5	0,4	0,7
25-34	84,2	80,2	88,1	84,5	83,9	85,0	84,4	83,9	85,0	15,5	11,3	19,3	14,4	13,9	15,0	14,4	13,9	15,0	0,3	-0,3	1,4	1,1	1,0	1,3	1,1	0,9	1,3
35-44	77,5	72,8	82,4	76,1	75,4	76,7	76,1	75,5	76,7	21,3	16,5	26,1	22,1	21,5	22,7	22,1	21,5	22,7	1,1	-0,2	2,5	1,8	1,6	2,0	1,8	1,6	2,0
45-54	69,0	61,9	76,1	62,2	61,4	63,0	62,3	61,5	63,0	27,7	21,0	35,0	33,9	33,1	34,7	33,8	33,1	34,6	3,3	0,2	5,8	3,9	3,6	4,3	3,9	3,6	4,2
55-64	45,2	37,7	52,7	44,9	44,1	45,8	44,9	44,1	45,8	48,4	40,5	55,6	47,6	46,8	48,5	47,6	46,8	48,5	6,4	2,8	10,6	7,4	7,0	7,9	7,4	7,0	7,9
65-74	18,8	11,9	26,4	29,1	28,3	29,9	29,0	28,1	29,8	53,5	44,4	62,5	56,8	55,9	57,7	56,8	55,9	57,7	27,7	19,2	35,6	14,0	13,4	14,7	14,2	13,6	14,8
>=75	13,2	7,0	18,3	16,1	15,3	16,8	16,0	15,3	16,8	44,2	35,4	51,9	56,5	55,5	57,5	56,3	55,3	57,3	42,6	34,3	50,7	27,4	26,5	28,3	27,7	26,8	28,6
Totale	71,9	70,0	73,9	65,6	65,3	65,9	65,7	65,3	66,2	21,8	19,9	23,6	28,5	28,3	28,8	28,4	28,1	28,7	6,3	5,2	7,4	5,9	5,7	6,0	5,9	5,7	6,0

Follow-up 2005 - Abbinabili e non abbinabili per classe di età, genere e disabilità (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Maschi																	
	Non disabile						Disabile											
	Non abbinabili		Abbinabili		Totale	Non abbinabili		Abbinabili		Totale								
	%	IC95%	%	IC95%	%	%	IC95%	%	IC95%	%	IC95%							
6-34	99,3	97,8	100,1	99,1	99,0	99,2	99,1	99,0	99,2	0,7	-0,3	1,4	0,9	0,7	1,0	0,9	0,7	1,0
35-64	97,2	95,4	99,1	98,6	98,4	98,7	98,6	98,4	98,7	2,8	0,8	4,7	1,4	1,3	1,6	1,4	1,3	1,6
65-69	100,0	100,0	100,0	95,7	94,9	96,4	95,7	94,9	96,5	-	-	-	4,3	3,5	5,1	4,3	3,5	5,1
70-74	69,4	53,7	86,3	92,7	91,6	93,7	92,3	91,3	93,5	30,6	13,5	46,5	7,3	6,6	8,8	7,7	6,6	8,8
75-79	75,3	55,4	100,2	86,7	85,0	88,2	86,6	85,0	88,2	24,7	-0,5	45,0	13,3	11,8	15,1	13,4	11,8	15,1
>=80	52,7	35,9	78,4	64,3	62,0	66,7	64,2	61,9	66,5	47,3	28,5	71,5	35,7	33,5	38,1	35,8	33,5	38,1
Totale	94,9	93,3	96,4	96,7	96,6	96,9	96,7	96,5	96,8	5,1	3,5	6,8	3,3	3,2	3,5	3,3	3,2	3,5

Classe di età	Femmine																	
	Non disabile						Disabile											
	Non abbinabili		Abbinabili		Totale	Non abbinabili		Abbinabili		Totale								
	%	IC95%	%	IC95%	%	%	IC95%	%	IC95%	%	IC95%							
6-34	99,2	98,0	100,1	99,2	99,0	99,3	99,2	99,0	99,3	0,8	-0,1	2,1	0,8	0,7	1,0	0,8	0,7	1,0
35-64	96,9	95,1	98,8	98,5	98,3	98,6	98,4	98,3	98,6	3,1	1,1	5,0	1,5	1,4	1,7	1,6	1,4	1,7
65-69	89,1	74,3	100,7	93,6	92,7	94,4	93,5	92,7	94,4	10,9	-1,0	26,0	6,4	5,5	7,3	6,5	5,6	7,4
70-74	81,9	67,4	94,5	88,7	87,5	89,8	88,6	87,4	89,7	18,1	5,2	32,9	11,3	10,1	12,5	11,4	10,2	12,6
75-79	68,1	53,2	82,8	79,4	77,8	81,0	79,2	77,6	80,7	31,9	17,1	46,9	20,6	19,0	22,3	20,8	19,2	22,4
>=80	31,1	19,1	42,5	51,6	49,8	53,3	51,1	49,4	52,8	68,9	57,6	80,8	48,4	46,7	50,2	48,9	47,2	50,6
Totale	90,5	88,6	92,5	93,9	93,7	94,1	93,9	93,7	94,1	9,5	7,6	11,7	6,1	5,9	6,3	6,1	5,9	6,3

Classe di età	Totale																	
	Non disabile						Disabile											
	Non abbinabili		Abbinabili		Totale	Non abbinabili		Abbinabili		Totale								
	%	IC95%	%	IC95%	%	%	IC95%	%	IC95%	%	IC95%							
6-34	99,2	98,6	99,9	99,1	99,0	99,2	99,1	99,1	99,2	0,8	0,1	1,5	0,9	0,8	1,0	0,9	0,8	1,0
35-64	97,1	95,8	98,4	98,5	98,4	98,6	98,5	98,4	98,6	2,9	1,5	4,3	1,5	1,4	1,6	1,5	1,4	1,6
65-69	93,9	86,6	100,5	94,6	94,0	95,1	94,5	94,0	95,1	6,1	-0,8	13,7	5,4	4,8	6,0	5,5	4,8	6,0
70-74	75,7	65,1	86,1	90,5	89,7	91,3	90,3	89,5	91,1	24,3	13,7	35,0	9,5	8,7	10,3	9,7	8,9	10,5
75-79	70,1	58,3	82,9	82,4	81,3	83,5	82,2	81,1	83,4	29,9	17,0	41,8	17,6	16,4	18,7	17,8	16,6	18,9
>=80	37,0	27,3	48,2	55,9	54,5	57,3	55,5	54,2	56,9	63,0	51,9	72,6	44,1	42,7	45,5	44,5	43,1	45,8
Totale	92,5	91,2	93,7	95,3	95,2	95,4	95,2	95,1	95,4	7,5	6,2	8,9	4,7	4,6	4,8	4,8	4,6	4,9

Follow-up 2005 - Abbinabili e non abbinabili per classe di età e abitudine al fumo (valori percentuali e intervalli di confidenza al 95%)

Classe di età	Abitudine al fumo																	
	Sì			No, ma fumava in passato						No, non ha mai fumato								
	Non abbinabili	Abbinabili	Totale	Non abbinabili	Abbinabili	Totale	Non abbinabili	Abbinabili	Totale	Non abbinabili	Abbinabili	Totale						
%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%					
14-24	22,2	19,5 24,9	22,2	21,3 23,0	22,2	21,3 23,0	4,7	3,4 6,1	5,4	4,9 5,8	5,3	4,9 5,7	73,1	70,3 76,2	72,5	71,5 73,4	72,5	71,6 73,5
25-44	30,5	28,4 32,6	33,1	32,5 33,8	32,9	32,3 33,6	14,7	13,1 16,3	15,6	15,1 16,2	15,6	15,1 16,1	54,8	52,5 57,1	51,2	50,5 51,9	51,5	50,8 52,2
45-64	27,2	24,9 29,5	26,4	25,8 27,1	26,5	25,8 27,2	21,6	19,5 23,8	24,6	24,0 25,3	24,4	23,8 25,0	51,2	48,6 53,8	48,9	48,1 49,7	49,1	48,3 49,8
>=65	9,2	7,6 10,9	10,9	10,4 11,5	10,8	10,2 11,3	24,5	22,0 26,9	28,4	27,6 29,2	28,0	27,3 28,8	66,3	63,6 69,0	60,7	59,8 61,6	61,2	60,3 62,0
Italia	23,5	22,4 24,7	25,1	24,7 25,4	24,9	24,6 25,3	17,1	16,1 18,1	19,3	19,0 19,7	19,1	18,8 19,5	59,4	58,0 60,8	55,6	55,2 56,1	55,9	55,5 56,3

Follow-up 2005 - Abbinabili e non abbinabili per classe di età e Body Mass Index (valori percentuali e intervalli di confidenza al 95%)

Classe di età	BMI in classi											
	Sottopeso						Normopeso					
	Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale	
%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	
18-24	6,9	3,0 11,3	9,8	9,2 10,5	9,8	9,1 10,4	72,7	65,5 79,4	75,5	74,6 76,5	75,5	74,5 76,4
25-34	5,7	3,0 8,2	5,9	5,5 6,3	5,9	5,5 6,3	70,6	65,5 75,3	67,8	67,1 68,6	67,9	67,2 68,6
35-44	4,3	1,7 6,4	2,9	2,7 3,2	3,0	2,7 3,2	66,8	61,2 72,1	58,8	58,1 59,6	59,0	58,3 59,7
45-54	1,1	-0,6 2,6	1,6	1,4 1,8	1,6	1,4 1,8	44,8	37,3 52,7	48,3	47,5 49,1	48,3	47,5 49,1
55-64	0,8	-0,6 2,5	1,2	1,0 1,4	1,2	1,0 1,4	41,7	34,8 49,8	40,8	40,0 41,7	40,8	40,0 41,7
65-74	2,8	-0,4 5,9	1,4	1,2 1,6	1,4	1,2 1,6	39,4	30,8 48,6	38,6	37,7 39,4	38,6	37,7 39,5
>=75	8,4	3,4 12,7	3,4	3,0 3,8	3,5	3,1 3,9	45,5	36,6 53,1	46,4	45,4 47,4	46,4	45,4 47,4
Totale	4,4	3,2 5,4	3,4	3,3 3,6	3,5	3,3 3,6	58,0	55,5 60,5	53,7	53,4 54,0	53,7	53,4 54,1

Classe di età	BMI in classi											
	Sovrappeso						Obeso					
	Non abbinabili		Abbinabili		Totale		Non abbinabili		Abbinabili		Totale	
%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	%	IC95%	
18-24	17,2	11,4 23,3	12,6	11,8 13,3	12,7	11,9 13,4	3,2	0,2 5,9	2,1	1,7 2,4	2,1	1,8 2,4
25-34	21,6	16,9 25,9	21,9	21,3 22,6	21,9	21,3 22,6	2,1	0,4 3,7	4,3	4,0 4,7	4,3	4,0 4,6
35-44	21,2	16,5 26,1	30,6	29,9 31,3	30,4	29,8 31,1	7,7	4,3 10,6	7,6	7,2 8,0	7,6	7,2 8,0
45-54	38,5	30,5 45,5	38,3	37,5 39,1	38,3	37,5 39,1	15,6	10,2 21,8	11,8	11,2 12,3	11,8	11,3 12,4
55-64	41,8	34,8 49,8	43,3	42,5 44,2	43,3	42,5 44,1	15,7	9,8 20,9	14,6	14,0 15,2	14,6	14,0 15,2
65-74	40,9	32,1 50,1	44,7	43,8 45,6	44,6	43,7 45,6	16,9	9,6 23,3	15,4	14,7 16,0	15,4	14,7 16,1
>=75	33,5	25,5 41,2	38,6	37,7 39,6	38,6	37,6 39,5	12,6	7,0 18,3	11,6	10,9 12,2	11,6	10,9 12,2
Totale	28,5	26,2 30,8	33,2	32,9 33,5	33,1	32,9 33,4	9,1	7,6 10,7	9,7	9,5 9,8	9,6	9,5 9,8

Appendice B

Rapporto decessi osservati/attesi per anno, sesso e classi di età - Follow-up 2000 e Follow-up 2005

Classe di età	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Totale
Follow-up 1999-2000															
Maschi															
0-14	0,74	0,54	0,34	0,00	0,42	0,50	1,81	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,45
15-24	0,91	0,79	1,45	0,66	0,95	0,76	1,52	0,28	1,18	0,68	0,00	0,38	2,21	0,47	0,89
25-34	0,48	1,28	0,68	0,88	0,77	0,80	0,86	0,36	1,40	0,43	1,15	0,96	1,04	1,41	0,86
35-44	0,29	1,21	0,69	0,80	0,83	1,19	1,24	1,12	0,88	0,96	0,99	0,86	1,40	0,85	0,93
45-54	0,74	0,67	1,02	0,65	0,96	1,14	0,84	0,85	0,98	0,97	0,79	1,26	1,24	0,76	0,91
55-64	0,99	1,03	0,90	0,86	0,82	1,14	1,17	0,85	0,95	1,13	0,90	1,01	1,05	1,32	1,00
65-74	0,90	0,90	0,84	1,07	0,89	1,03	1,06	0,89	1,11	1,02	0,89	0,97	1,00	1,22	0,98
75+	0,83	0,92	0,91	0,96	0,86	0,87	0,93	0,90	0,83	0,89	0,98	0,85	1,00	0,94	0,91
Totale	0,85	0,92	0,89	0,95	0,86	0,95	0,99	0,89	0,91	0,93	0,95	0,90	1,02	1,01	0,93
Femmine															
0-14	0,00	0,00	0,00	0,00	1,68	0,00	0,00	0,00	1,05	0,00	0,00	0,00	0,00	0,00	0,21
15-24	0,00	0,00	0,00	1,51	0,80	1,56	0,00	0,90	2,58	0,00	0,99	1,16	0,00	1,31	0,73
25-34	2,30	0,68	0,00	0,78	0,43	0,91	1,02	0,51	0,53	0,59	1,18	1,23	1,88	0,72	0,91
35-44	1,20	0,82	1,36	1,33	1,44	1,10	0,79	0,84	1,21	0,71	0,92	0,93	0,78	0,83	1,04
45-54	0,76	1,07	0,95	0,89	0,83	1,06	0,92	0,79	0,81	1,55	0,88	1,07	0,98	1,09	0,97
55-64	1,07	0,78	0,96	1,04	0,76	0,87	1,08	0,98	1,07	1,09	0,82	0,97	1,01	0,83	0,95
65-74	0,85	1,09	0,95	1,09	1,11	1,13	1,02	0,97	0,99	0,90	1,06	0,92	0,89	0,93	0,99
75+	0,76	0,81	0,82	0,87	0,82	0,89	0,85	0,95	0,83	0,95	0,84	0,78	0,87	0,88	0,85
Totale	0,80	0,85	0,85	0,92	0,86	0,92	0,88	0,95	0,87	0,96	0,87	0,81	0,88	0,89	0,88
Follow-up 2004-2005															
Maschi															
0-14						0,00	0,72	0,39	0,43	0,00	0,60	1,47	0,00	0,00	0,39
15-24						1,05	0,53	0,57	0,89	0,69	0,34	1,15	0,87	0,47	0,73
25-34						1,18	0,72	1,16	1,30	0,00	0,25	1,54	1,10	1,49	0,96
35-44						0,77	0,70	0,65	0,58	0,74	0,55	0,33	1,08	1,13	0,71
45-54						0,67	0,68	0,92	0,64	0,70	1,09	0,80	1,25	0,72	0,83
55-64						0,78	1,12	0,95	0,89	1,09	0,83	0,92	0,77	0,89	0,92
65-74						0,89	0,80	0,88	0,79	0,98	0,94	0,83	0,95	0,91	0,89
75+						0,98	0,91	0,89	0,94	0,88	0,90	1,00	0,95	0,91	0,93
Totale						0,92	0,89	0,89	0,89	0,90	0,90	0,95	0,94	0,91	0,91
Femmine															
0-14						0,86	0,91	0,00	0,00	0,00	0,76	0,00	1,20	1,41	0,51
15-24						1,63	2,51	0,00	0,86	0,00	0,97	1,11	2,22	0,00	1,08
25-34						0,46	1,02	1,05	1,08	1,82	1,21	3,79	0,65	0,74	1,28
35-44						0,62	0,78	1,67	0,52	0,53	0,73	1,31	0,39	1,86	0,92
45-54						0,54	0,54	0,60	1,08	0,79	1,02	0,76	0,59	1,23	0,80
55-64						0,81	1,07	0,67	1,18	0,89	1,13	1,11	0,97	1,23	1,01
65-74						0,99	1,13	1,02	0,94	0,84	1,09	0,98	0,80	0,81	0,96
75+						0,89	0,84	0,87	0,89	0,89	0,79	0,90	0,89	0,89	0,87
Totale						0,89	0,88	0,88	0,91	0,88	0,84	0,93	0,87	0,91	0,89

Appendice C

Rapporto ricoveri osservati nel campione su ricoveri attesi per anno, sesso e classe di età -Follow-up 2000

Rapporto osservati/attesi Anno 2001 (a)

Classi di età	Maschi	Femmine	Totale
1-14	69,94	73,95	71,69
15-24	72,35	72,66	72,52
25-34	80,97	73,38	76,25
35-44	78,88	73,27	75,67
45-54	75,12	75,75	75,42
55-64	79,58	79,77	79,75
65-74	77,89	75,92	77,53
75+	74,06	67,69	71,3
Totale	76,37	73,68	75,29

(a) esclusi Sardegna e minori di 1 anno

Rapporto osservati/attesi Anno 2002 (a)

Classi di età	Maschi	Femmine	Totale
2-14	69,39	66,18	68,07
15-24	83,18	76,23	79,41
25-34	79,54	72,76	75,16
35-44	77,12	72,38	74,46
45-54	80,08	74,63	77,19
55-64	83,84	82,67	83,38
65-74	80,34	81,26	81,19
75+	78,74	73,54	76,75
Totale	79,52	75,61	77,79

(a) esclusi Sardegna e minori di 2 anni

Rapporto osservati/attesi Anno 2003

Classi di età	Maschi	Femmine	Totale
3-14	72,69	65,66	69,72
15-24	83,50	78,38	80,72
25-34	82,42	81,15	81,74
35-44	80,15	85,16	83,73
45-54	82,65	81,12	81,80
55-64	87,52	88,50	88,03
65-74	83,70	87,86	86,03
75+	79,94	77,84	79,55
Totale	82,22	82,10	82,48

(a) esclusi minori di 3 anni

Rapporto osservati/attesi Anno 2004

Classi di età	Maschi	Femmine	Totale
4-14	72,78	68,72	71,07
15-24	84,30	82,00	83,02
25-34	79,54	78,35	78,88
35-44	86,22	89,12	88,55
45-54	86,80	89,28	88,04
55-64	86,91	87,58	87,30
65-74	81,11	84,83	83,12
75+	83,26	82,16	83,45
Totale	83,24	83,98	83,94

(a) esclusi minori di 4 anni

Rapporto osservati/attesi Anno 2005

Classi di età	Maschi	Femmine	Totale
5-14	83,22	85,29	84,10
15-24	84,27	73,91	78,52
25-34	87,18	81,13	82,99
35-44	71,97	85,33	80,91
45-54	89,85	87,28	88,47
55-64	84,80	82,78	83,94
65-74	87,24	89,80	88,75
75+	81,57	82,30	82,75
Totale	84,15	84,03	84,40

(a) esclusi minori di 5 anni

Rapporto osservati/attesi Anno 2006

Classi di età	Maschi	Femmine	Totale
6-14	86,15	71,55	79,74
15-24	78,14	77,64	77,8
25-34	83,76	82,68	82,79
35-44	77,18	83,73	81,93
45-54	91,98	88,58	90,17
55-64	85,31	83,97	84,78
65-74	89,36	85,03	87,78
75+	83,42	83,39	84,2
Totale	85,42	83,58	84,75

(a) esclusi minori di 6 anni

Rapporto osservati/attesi Anno 2007

Classi di età	Maschi	Femmine	Totale
7-14	85,63	68,14	77,92
15-24	87,23	69,04	77,13
25-34	81,47	77,60	78,48
35-44	85,42	90,61	89,46
45-54	87,34	85,81	86,51
55-64	87,05	91,19	89,05
65-74	89,07	86,37	88,20
75+	85,13	87,68	87,33
Totale	86,56	85,18	86,16

(a) esclusi minori di 7 anni

Rapporto osservati/attesi Anno 2008

Classi di età	Maschi	Femmine	Totale
8-14	79,02	68,10	74,12
15-24	82,20	70,71	75,77
25-34	93,62	83,37	85,82
35-44	82,45	89,37	87,57
45-54	89,58	82,55	85,87
55-64	88,52	94,30	91,25
65-74	88,20	91,46	89,94
75+	86,15	87,57	87,72
Totale	87,12	86,82	87,25

(a) esclusi minori di 8 anni

Rapporto osservati/attesi Anno 2009

Classi di età	Maschi	Femmine	Totale
9-14	87,95	69,95	79,85
15-24	80,70	79,27	79,73
25-34	88,89	81,67	83,02
35-44	78,11	87,71	84,95
45-54	82,93	85,87	84,49
55-64	92,49	93,90	93,24
65-74	89,43	87,11	88,69
75+	87,23	84,11	86,47
Totale	87,22	85,70	86,68

(a) esclusi minori di 9 anni

Rapporto osservati/attesi Anno 2010

Classi di età	Maschi	Femmine	Totale
10-14	76,44	58,15	68,39
15-24	101,64	75,44	87,10
25-34	89,76	82,47	83,79
35-44	88,30	92,62	91,51
45-54	86,39	91,23	88,95
55-64	88,92	92,67	90,76
65-74	87,40	89,50	88,52
75+	89,60	86,68	88,94
Totale	88,85	87,63	88,45

(a) esclusi minori di 10 anni

Rapporto osservati/attesi Anno 2011

Classi di età	Maschi	Femmine	Totale
11-14	78,98	56,08	68,73
15-24	94,62	80,87	86,91
25-34	95,91	82,22	85,23
35-44	79,93	96,30	91,14
45-54	89,70	96,38	93,23
55-64	89,50	89,33	89,60
65-74	86,90	93,57	90,04
75+	82,40	84,00	84,00
Totale	86,54	88,52	87,83

(a) esclusi minori di 11 anni

Rapporto osservati/attesi Anno 2012

Classi di età	Maschi	Femmine	Totale
12-14	55,04	35,97	46,69
15-24	94,79	70,68	81,32
25-34	94,17	82,66	84,95
35-44	92,66	90,31	91,21
45-54	88,33	93,71	91,20
55-64	88,84	92,41	90,62
65-74	89,59	88,31	89,23
75+	86,28	83,61	85,60
Totale	88,58	86,27	87,56

(a) esclusi minori di 12 anni

Rapporto osservati/attesi Anno 2013

Classi di età	Maschi	Femmine	Totale
13-14	21,32	26,94	23,72
15-24	91,03	77,43	83,42
25-34	88,16	86,97	86,47
35-44	91,80	89,84	90,47
45-54	87,39	89,81	88,70
55-64	92,07	96,22	94,14
65-74	92,35	94,48	93,42
75+	93,81	86,59	90,77
Totale	91,03	88,89	90,08

(a) esclusi minori di 13 anni

Rapporto osservati/attesi Anno 2014

Classi di età	Maschi	Femmine	Totale
14	25,98	15,73	21,28
15-24	87,16	77,52	81,80
25-34	92,99	85,41	86,66
35-44	93,00	103,24	99,86
45-54	91,53	86,09	88,67
55-64	91,45	92,68	92,24
65-74	93,04	88,88	91,38
75+	91,72	91,30	92,02
Totale	91,78	90,49	91,26

(a) esclusi minori di 14 anni

Rapporto ricoveri osservati nel campione su ricoveri attesi per anno, sesso e classe di età - Follow-up 2005
Rapporto osservati/attesi Anno 2006 (a)

Classi di età	Maschi	Femmine	Totale
1-14	77,69	78,30	77,97
15-24	91,17	75,65	82,52
25-34	87,83	79,52	82,35
35-44	85,27	90,34	89,00
45-54	88,66	88,78	88,73
55-64	87,74	89,46	88,57
65-74	83,11	86,63	84,76
75+	89,55	85,29	87,85
Totale	86,35	85,36	86,05

(a) esclusi minori di 1 anno

Rapporto osservati/attesi Anno 2007 (a)

Classi di età	Maschi	Femmine	Totale
2-14	80,06	80,56	80,29
15-24	86,50	71,29	78,02
25-34	87,72	80,36	82,75
35-44	82,34	88,84	87,04
45-54	89,89	85,10	87,37
55-64	84,88	91,65	88,03
65-74	86,15	92,37	88,99
75+	88,74	83,73	86,67
Totale	86,28	85,62	86,15

(a) esclusi minori di 2 anni

Rapporto osservati/attesi Anno 2008 (a)

Classi di età	Maschi	Femmine	Totale
3-14	75,76	71,56	73,97
15-24	88,04	76,25	81,42
25-34	88,78	79,07	82,03
35-44	83,28	88,50	87,04
45-54	88,19	90,16	89,26
55-64	93,31	93,28	93,34
65-74	89,77	89,89	89,95
75+	93,92	88,71	91,64
Totale	89,58	86,91	88,36

(a) esclusi minori di 3 anni

Rapporto osservati/attesi Anno 2009 (a)

Classi di età	Maschi	Femmine	Totale
4-14	78,12	81,69	79,67
15-24	86,23	66,96	75,56
25-34	84,21	81,38	82,20
35-44	80,88	92,04	88,52
45-54	84,35	90,15	87,44
55-64	90,14	90,85	90,43
65-74	88,88	87,94	88,67
75+	90,85	89,57	90,62
Totale	87,41	87,25	87,50

(a) esclusi minori di 4 anni

Rapporto osservati/attesi Anno 2010 (a)

Classi di età	Maschi	Femmine	Totale
5-14	72,53	86,93	78,66
15-24	88,33	72,81	79,65
25-34	88,51	81,34	83,07
35-44	79,59	94,84	89,94
45-54	85,56	92,08	89,03
55-64	86,15	91,00	88,35
65-74	89,75	94,04	91,98
75+	92,86	90,36	91,87
Totale	87,67	89,57	88,82

(a) esclusi minori di 5 anni

Rapporto osservati/attesi Anno 2011 (a)

Classi di età	Maschi	Femmine	Totale
6-14	79,65	81,37	80,41
15-24	96,80	73,51	83,89
25-34	94,03	82,55	85,28
35-44	87,36	92,21	90,91
45-54	86,99	91,34	89,29
55-64	90,46	89,81	90,18
65-74	99,33	89,59	95,34
75+	96,44	87,31	91,97
Totale	93,45	87,54	90,45

(a) esclusi minori di 6 anni

Rapporto osservati/attesi Anno 2012 (a)

Classi di età	Maschi	Femmine	Totale
7-14	78,45	77,31	78,03
15-24	95,79	76,05	84,74
25-34	90,77	80,49	82,68
35-44	101,12	92,40	95,40
45-54	89,67	91,81	90,84
55-64	91,23	90,42	90,92
65-74	95,03	89,15	92,77
75+	94,77	92,39	93,78
Totale	93,34	88,85	91,05

(a) esclusi minori di 7 anni

Rapporto osservati/attesi Anno 2013 (a)

Classi di età	Maschi	Femmine	Totale
8-14	80,82	61,82	72,85
15-24	91,52	69,36	79,14
25-34	98,95	81,21	85,09
35-44	95,96	96,53	96,62
45-54	87,27	87,95	87,66
55-64	84,73	98,03	90,83
65-74	93,04	85,88	90,23
75+	94,06	88,51	91,29
Totale	91,30	87,70	89,47

(a) esclusi minori di 8 anni

Rapporto osservati/attesi Anno 2014 (a)

Classi di età	Maschi	Femmine	Totale
9-14	76,53	77,02	76,84
15-24	90,29	75,30	81,99
25-34	88,54	82,58	83,26
35-44	93,11	95,47	94,93
45-54	99,75	93,58	96,49
55-64	92,30	96,80	94,43
65-74	94,58	91,49	93,45
75+	94,45	94,56	94,71
Totale	93,62	91,57	92,60

(a) esclusi minori di 9 anni

Riferimenti bibliografici

- Caranci, N., B. Pacelli e A. Ranzi. 2014. Un sistema per il monitoraggio delle disuguaglianze di salute e per la valutazione delle azioni di contrasto. *L'Equità nella salute in Italia* (a cura di Costa G. et al);224-257. Milano. Franco Angeli Editore.
- Caranci, N., C. Di Girolamo, P. Giorgi Rossi, T. Spadea, B. Pacelli, S. Broccoli, P. Ballottari, G. Costa, N. Zengarini, N. Agabiti, A. M. Bargagli, L. Cacciani, C. Canova, L. Cestari, A. Biggeri, L. Grisotto, G. Terni, G. Costanzo, C. Mirisola e A. Petrelli on behalf of the IN-LiMeS Group. 2018. Cohort profile: the Italian Network of Longitudinal Metropolitan Studies (IN-LiMeS), a multicentre cohort for socioeconomic inequalities in health monitoring. *BMJ Open* 2018;8:e020572.
- Costa, G.,M. Bassi, G.F. Gensini, M. Marra, A.L. Nicelli e N. Zengarini (a c. di). 2014. *L'Equità nella salute in Italia*. Milano. Franco Angeli Editore.
- D'Errico, A., P. Carnà, C. Piccinelli, G. Sebastiani e G. Costa. 2016. Disoccupazione e mortalità. Salute in Italia e livelli di tutela: approfondimenti dalle indagini ISTAT sulla salute (a cura di Costa G. et altri). *Rapporti ISTISAN* 16|26. (ISSN:1123-3117 (cartaceo) - 2384-8936 (online), 225-231.
- Eikemo, T.A., R. Hoffmann, M. C. Kulik, I. Kulhánová, M. Toch-Marquardt, G. Menvielle, C. Looman, D. Jasilionis, P. Martikainen, O. Lundberg e J. P. Mackenbach and for the EURO-GBD-SE consortium. 2014. How Can Inequalities in Mortality Be Reduced? A Quantitative Analysis of 6 Risk Factors in 21 European Populations. Published in *PloS one*. DOI:10.1371/journal.pone.0110952
- Federico, B., J.P. Mackenbach, T.A. Eikemo, G. Sebastiani, C. Marinacci, G. Costa e A. E. Kunst. 2013. Educational inequalities in mortality in northern, mid and southern Italy and the contribution of smoking. *Journal of Epidemiology and Community Health* 2013;67:603–609. doi:10.1136/jech-2012-201716, 17.
- Mackenbach, J.P. 2005. *Health inequalities. Europe in profile*. UK Presidency of the EU, London 2005.
- Mackenbach, J.P., Stirbu I., Roskam A.J., Schaap M., Menvielle G., Leinsalu M. e Kunst A. 2008. Socioeconomic inequalities in health in 22 European countries. *New England Journal of Medicine*; 358:2468–81.
- Mackenbach, J.P., I. Kulhánová, B. Artnik, M. Bopp, C. Borrell e T. Clemens, et al. 2016. Changes in mortality inequalities over two decades: register based study of European countries. *BMJ*, doi: 10.1136/bmj.i1732.
- Mäki, N.,P. Martikainen, T.A. Eikemo, G. Menvielle, O. Lundberg, O. Ostergren, D. Jasilionis e J.P. Mackenbach, EURO-GBD-SE consortium. 2011. Educational differences in disability-free life expectancy: a comparative study of long-standing activity limitation in eight European countries in the EURO-GBD-SE project.
- Marinacci, C., F. Grippo, M. Pappagallo, G. Sebastiani, M. Demaria, P. Vittori, N. Caranci e G. Costa. 2013. Social inequalities in total and cause-specific mortality of a sample of the Italian population, from 1999 to 2007. *European Journal of Public Health*, Vol. 23, No. 4, 582–587.
- Marinacci, C., G. Sebastiani, E. Ferracin, F. Grippo e M. Pappagallo, T. Landriscina, L. Gargiulo, M. Demaria e G. Costa. 2011. La distanza tra gli Italiani attraverso le disuguaglianze geografiche e socioeconomiche. *Epidemiologia e Prevenzione*, anno 35 (5) settembre-ottobre Suppl. 1
- Marmot, M. 2005. Social determinants of health inequalities. *Lancet.*; 365:1099–1104.

- Piccinelli, C., P. Carnà, S. Stringhini, Sebastiani G., M. Demaria, M. Marra, G. Costa e A. D'Errico 2018. The contribution of behavioural and metabolic risk factors to socioeconomic inequalities in mortality: the Italian Longitudinal Study. *International Journal of Public Health* <https://doi.org/10.1007/s00038-018-1076-8>.
- Piccinelli, C., P. Carnà, G. Costa, L. Gargiulo, G. Sebastiani e A. D'Errico. Fumo, sedentarietà e sovrappeso in relazione alle disuguaglianze nella mortalità in Italia: risultati dallo studio longitudinale italiano. 2016. *Salute in Italia e livelli di tutela: approfondimenti dalle indagini ISTAT sulla salute (a cura di Costa G. et al) . – Rapporti ISTISAN 16|26 . (ISSN:1123-3117 (cartaceo) - 2384-8936 (online), 94-105.*
- Stringhini, S., T. Spadea, M. Stroschia, R. Onorati, M. Demaria e N. Zengarini, et al. 2015. Decreasing educational differences in mortality over 40 years: evidence from the Turin Longitudinal Study (Italy). *Journal of Epidemiology & Community Health*, 69:1208–1216.

Informazioni per gli autori

La collana è aperta alle autrici e agli autori dell'Istat e del Sistema statistico nazionale e ad altri studiosi che abbiano partecipato ad attività promosse dall'Istat, dal Sistan, da altri Enti di ricerca e dalle Università (convegni, seminari, gruppi di lavoro, ecc.).

Coloro che desiderano pubblicare su questa collana devono sottoporre il proprio contributo al Comitato di redazione degli *Istat working papers*, inviandolo per posta elettronica all'indirizzo: iwp@istat.it.

Il saggio deve essere redatto seguendo gli standard editoriali previsti (disponibili sul sito dell'Istat), corredato di un sommario in Italiano e in Inglese e accompagnato da una dichiarazione di paternità dell'opera.

Per le autrici e gli autori dell'Istat, la sottomissione dei lavori deve essere accompagnata da un'e-mail della/del propria/o referente (Direttrice/e, Responsabile di Servizio, etc.), che ne assicura la presa visione.

Per le autrici e gli autori degli altri Enti del Sistan la trasmissione avviene attraverso la/il responsabile dell'Ufficio di statistica, che ne prende visione. Per tutte le altre autrici e gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione.

Per la stesura del testo occorre seguire le indicazioni presenti nel foglio di stile, con le citazioni e i riferimenti bibliografici redatti secondo il protocollo internazionale 'Autore-Data' del *Chicago Manual of Style*.

Attraverso il Comitato di redazione, tutti i lavori saranno sottoposti a un processo di valutazione doppio e anonimo che determinerà la significatività del lavoro per il progresso dell'attività statistica istituzionale.

La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line gratuitamente.

Gli articoli pubblicati impegnano esclusivamente le autrici e gli autori e le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat.

Si autorizza la riproduzione a fini non commerciali e con citazione della fonte.