

Protocollo di ricerca

Rafforzare la programmazione sanitaria regionale attraverso l'analisi integrata dei dati amministrativi

Diagnosi, sopravvivenza e cause di morte: evidenze per il decision making

Dr.ssa Benedetta Armocida MD PhD

ISS – Dipartimento Malattie Cardiovascolari, Endocrino Metaboliche e dell'Invecchiamento (MACA)

Dr. Lorenzo Monasta DSc MSc

IRCCS Burlo Garofolo – Coordinatore Italian GBD Initiative

Autori e Collaboratori

Autori principali: Benedetta Armocida (ISS-MACA), Lorenzo Monasta (IRCCS Burlo Garofolo)

Coordinamento: Benedetta Armocida (ISS-MACA), Lorenzo Monasta (IRCCS Burlo Garofolo), Lisa Leonardini (ProMIS)

Responsabile del team IHME: Mohsen Naghavi (IHME)

Responsabile della ricerca IHME: Eve Wool (IHME)

Staff tecnico IHME: Vincent Mouglin, Authia Gray, Rebecca Hsu

Indice

Autori e Collaboratori	2
Indice	3
Abbreviazioni	4
1. Obiettivi del progetto di ricerca	5
2. Background e contesto internazionale	5
3. Scopo	6
4. Metodologia	6
4.1 Accesso ai dati	6
4.2 Pulizia dei dati	7
4.3 Pipeline Analitica	7
4.4 Metriche e analisi	8
5. Esiti attesi e significato	8
6. Sostenibilità	9
7. Aspetti etici e protezione dei dati	9
8. Team di ricerca e coordinamento	10
9. Regioni partecipanti	10
10. Documenti e file da produrre	10
Bibliografia	11

Abbreviazioni

Abbreviazione	Definizione	Informazioni / link
CSMR	Cause-Specific Mortality Rate	Tasso di mortalità causa-specifico
DCA	Data Collection Agency	Organizzazione incaricata della raccolta dati
GDPR	General Data Protection Regulation	Regolamento (UE) 2016/679
GBD	Global Burden of Disease	Studio epidemiologico globale coordinato da IHME
ICD-10	International Classification of Diseases, 10th Revision	Classificazione internazionale delle malattie
IHME	Institute for Health Metrics and Evaluation	https://www.healthdata.org
ISS	Istituto Superiore di Sanità	Ente coordinatore nazionale
IRCCS	Istituto di Ricovero e Cura a Carattere Scientifico	
Italian GBD Initiative	Iniziativa italiana nell'ambito del GBD	Coordinata dall'IRCCS Burlo Garofolo
MACA	Dipartimento Malattie Cardiovascolari, Endocrino Metaboliche e dell'Invecchiamento	Dipartimento ISS coinvolto nel progetto
NSC	National Study Center	ISS-MACA per il progetto italiano
ProMIS	Programma Mattone Internazionale Salute	
SDO	Scheda di Dimissione Ospedaliera	Fonte dati ospedalieri regionali
SSN	Servizio Sanitario Nazionale	

1. Obiettivi del progetto di ricerca

Il progetto si inserisce nel quadro del Global Burden of Disease Study (GBD), coordinato dall'Institute for Health Metrics and Evaluation (IHME) dell'Università di Washington, e mira a trasferire e adattare l'esperienza metodologica internazionale al contesto italiano. Il progetto è promosso dall'ISS – Dipartimento MACA, dal ProMIS – Programma Mattone Internazionale Salute e dall'IRCCS Burlo Garofolo in qualità di coordinatore dell'Italian GBD Initiative.

L'obiettivo generale è valorizzare i dati amministrativi regionali per produrre evidenze avanzate sul rapporto tra storia diagnostica individuale, determinanti sociali della salute e mortalità, a supporto della programmazione sanitaria regionale e nazionale.

I seguenti ambiti di analisi costituiscono il nucleo del progetto:

- Analisi di sopravvivenza e calcolo del rischio relativo e dell'eccesso di mortalità per ogni diagnosi presente nei dati amministrativi italiani.
- Inclusione dei determinanti sociali della salute (livello di istruzione, reddito) e stima della perdita di capitale umano.
- Costruzione di uno strumento di visualizzazione interattivo per presentare i risultati e stimolare nuove ricerche.
- Confronto internazionale dei risultati italiani con i Paesi già partecipanti (Brasile, Colombia, Svezia) e futuri partecipanti (Norvegia, Danimarca).

2. Background e contesto internazionale

Il Global Burden of Disease Study (GBD) rappresenta il più ampio e strutturato sforzo di collaborazione scientifica globale per la stima comparabile del carico di malattia, della mortalità e dei fattori di rischio, coinvolgendo oltre 18.000 ricercatori in più di 160 Paesi. L'IHME coordina il network globale di collaboratori del GBD, fornendo metodologie armonizzate, strumenti analitici avanzati e un'infrastruttura condivisa per il rafforzamento dei sistemi informativi sanitari nazionali.

In Italia, l'Italian GBD Initiative, coordinata dall'IRCCS Burlo Garofolo, rappresenta un esempio consolidato di integrazione tra competenze nazionali e network scientifici internazionali, con l'obiettivo di migliorare la qualità, l'uso e l'interpretazione dei dati di sanità pubblica. Questa collaborazione ha consentito la generazione di stime per l'Italia anche a livello regionale (19 regioni e due province autonome).

In questo contesto, la collaborazione con IHME si inserisce pienamente negli obiettivi del ProMIS – Programma Mattone Internazionale Salute, rafforzando i processi di internazionalizzazione del Servizio Sanitario Nazionale attraverso:

- l'adozione di metodologie condivise a livello globale;
- il confronto sistematico con Paesi dotati di sistemi informativi avanzati;
- il trasferimento di competenze analitiche e strumenti innovativi verso le istituzioni italiane.

Dal 2020, IHME collabora con istituzioni accademiche e governative per analizzare il legame tra la storia diagnostica individuale, i determinanti sociali della salute e le cause di morte, utilizzando dati amministrativi sanitari linkati ai registri di mortalità. Questo approccio consente di:

- stimare il rischio relativo di morte associato a specifiche diagnosi;
- stimare l'eccesso di mortalità attribuibile a ciascuna condizione nella popolazione, fornendo evidenze cruciali per la programmazione sanitaria e la valutazione dell'impatto delle politiche di prevenzione e cura.

Utilizzando una pipeline di codice ottimizzata (*optimized code pipeline*) che sfrutta i metodi di elaborazione delle cause di morte dell'IHME, queste metriche vengono create per ogni combinazione di diagnosi e causa di morte. In questo modo, siamo in grado di integrare le diagnosi ricevute durante le visite ambulatoriali e i ricoveri ospedalieri. Questi risultati vengono quindi rappresentati in uno strumento di visualizzazione che può stimolare ulteriori ricerche su combinazioni diagnosi-causa uniche, far comprendere il rischio di ricovero dovuto a una malattia, fungere da evidenza per proposte di ricerca e finanziamenti e diventare una base di confronto con altri paesi partecipanti.

I ricercatori dell'IHME stanno attualmente lavorando in collaborazione con ricercatori dell'Università Federale di Minas Gerais, del Karolinska Institutet, dell'Istituto Norvegese di Salute Pubblica, dell'Istituto Statistico Danese e dell'Università de La Sabana per condurre questa ricerca con dati rappresentativi a livello nazionale, rispettivamente, di Brasile, Svezia, Norvegia, Danimarca e Colombia. Di questi progetti, è stata completata la produzione di risultati preliminari per Brasile, Colombia e Svezia.

La ricerca sul Database Nazionale Brasiliano della Salute ha prodotto un set di dati linkati comprendente 72 milioni di individui, oltre 165 milioni di ricoveri ospedalieri e 5,1 milioni di decessi dal 2000 al 2015. Questo lavoro è stato alla base dello studio di Souto Melo et al., pubblicato nel 2022, che ha valutato l'effetto delle gravi malattie mentali sulla mortalità per tutte le cause e causa-specifica nella popolazione di pazienti brasiliani [1]. Il set di dati linkati dai registri svedesi comprende 11 milioni di individui, 2 milioni di decessi e oltre 500 milioni di diagnosi ospedaliere. In collaborazione con il Karolinska Institutet, quest'area di ricerca viene ampliata per includere la stratificazione tra determinanti sociali della salute, come il livello di istruzione e il reddito, e per costruire tavole di mortalità e diagrammi di Lexis che sfruttano la costruzione di coorti per stimare la sopravvivenza comparativa tra popolazioni diagnosticate e non diagnosticate.

Il presente progetto mira a trasferire e adattare questa esperienza internazionale al contesto italiano, valorizzando i dati amministrativi regionali e rafforzando la capacità nazionale di analisi avanzata, in linea con le priorità europee di cooperazione scientifica e con iniziative quali le COST Actions e le Joint Actions europee in ambito NCDs, prevenzione e sistemi informativi sanitari.

3. Scopo

Il presente protocollo definisce gli standard e le procedure per la conduzione del progetto di ricerca. In particolare, il progetto si prefigge di:

1. Eseguire analisi di sopravvivenza e calcolare il rischio relativo e l'eccesso di mortalità per ogni diagnosi presente nei dati amministrativi italiani tenendo conto dei determinanti sociali della salute e della perdita di capitale umano.
2. Costruire uno strumento di visualizzazione completo per presentare i risultati e generare nuovi stimoli di ricerca.
3. Confrontare i risultati dell'Italia con quelli dei Paesi attualmente partecipanti (Brasile, Colombia, Svezia) e dei Paesi futuri partecipanti (Norvegia, Danimarca), rafforzando l'integrazione dell'Italia nel network internazionale del Global Burden of Disease Study e contribuendo al miglioramento dei modelli GBD a livello globale.

4. Metodologia

4.1 Accesso ai dati

L'IHME, il ProMIS e l'ISS-MACA creeranno congiuntamente un piano di lavoro per consentire il corretto svolgimento delle attività. Vincent Mougín, Authia Gray e Rebecca Hsu sono i membri dello staff dell'IHME

che lavoreranno sui dati. Mohsen Naghavi è il responsabile del team ed Eve Wool è la responsabile della ricerca. I dati non saranno trasferiti ai server dell'IHME.

Il NSC/NST deve tenere conto di quanto segue nell'impostazione del progetto a livello regionale:

- Conformità al Regolamento generale sulla protezione dei dati (GDPR) e alla normativa nazionale applicabile in materia di protezione dei dati.
- Conformità agli standard etici applicabili e alle normative regionali vigenti.
- Identificazione di referenti tecnici regionali per la governance del progetto.
- Condivisione delle priorità analitiche locali con il coordinamento nazionale.

4.2 Pulizia dei dati

Utilizzando identificatori univoci dei pazienti, i dati verranno ripuliti in modo che vengano conservate le seguenti variabili: data di nascita, età, sesso, Regione di residenza, variabili relative all'istruzione (se disponibili), variabili relative al reddito (se disponibili), data della visita ambulatoriale, data della visita ospedaliera e/o del ricovero, diagnosi di ogni ricovero/visita, data del decesso (ove applicabile), causa principale del decesso (ove applicabile), cause multiple del decesso (ove applicabile).

Le diagnosi saranno quindi mappate dai codici ICD-10 al sistema di codifica delle cause di mortalità utilizzato nel Global Burden of Disease. Le cause primarie di morte che risultano troppo generiche, impossibili da identificare o di morte intermedie/immediate saranno ridistribuite tra cause di morte ben definite, seguendo la metodologia di redistribuzione adottata nel GBD.

4.3 Pipeline Analitica

La metodologia si basa sulla pipeline analitica ottimizzata dell'IHME, articolata nelle seguenti fasi:

Fase	Descrizione
Linkage	Collegamento dei record ospedalieri e ambulatoriali ai record di decesso tramite identificatori univoci pseudonimizzati.
Mapping	Mappatura dei codici diagnostici e delle cause di morte ai sistemi di codifica GBD (a_cause, b_cause, sindromi e patogeni infettivi). Utilizzo delle mappe fatali e non fatali delle cause del progetto.
Ridistribuzione	Applicazione degli algoritmi di redistribuzione GBD per correggere cause di morte mal classificate o generiche (garbage codes). Ridistribuzione coerente con i metodi GBD.
Aggregazione	Aggregazione dei dati diagnostici e di mortalità al livello demografico e diagnostico appropriato.
Matching e Coorti	Generazione di coorti di individui non diagnosticati per causa GBD di Livello 3, con composizione demografica identica alla coorte diagnosticata. Costruzione delle tavole di mortalità e diagrammi di Lexis espandendo i dati in formato person-time per: tempo dalla diagnosi o causa di morte, età, anno di calendario, sesso.

La pipeline contempla inoltre l'integrazione con i dati sui determinanti sociali della salute (reddito, istruzione) a livello individuale e per anno di diagnosi, nonché la stima della perdita di capitale umano.

4.4 Metriche e analisi

Le metriche principali calcolate per ogni coppia diagnosi/causa di morte sono le seguenti:

Analisi di sopravvivenza	Analisi dei diagrammi di Lexis, curve di sopravvivenza basate sulle coorti Lexis, diagrammi di Kaplan-Meier e modelli di rischio proporzionale di Cox. Stima della sopravvivenza per ogni patologia nell'arco di oltre 20 anni.																
		<table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Deceduto per causa Y</th> </tr> <tr> <th colspan="2"></th> <th>Sì</th> <th>No</th> </tr> </thead> <tbody> <tr> <th rowspan="2">Diagnosticato con X</th> <th>Sì</th> <td style="background-color: #cccccc;">A</td> <td style="background-color: #cccccc;">B</td> </tr> <tr> <th>No</th> <td style="background-color: #ffcccc;">C</td> <td style="background-color: #ccffff;">D</td> </tr> </tbody> </table>				Deceduto per causa Y				Sì	No	Diagnosticato con X	Sì	A	B	No	C
		Deceduto per causa Y															
		Sì	No														
Diagnosticato con X	Sì	A	B														
	No	C	D														
Rischio Relativo (RR)	Rapporto tra il tasso di mortalità causa-specifico (CSMR) della popolazione diagnosticata e quello della popolazione non diagnosticata:																
	$\frac{CSMR_{diagnosticato}}{CSMR_{non-diagnosticato}} = \frac{\frac{A}{(A+B)}}{\frac{C}{(C+D)}}$																
Eccesso di Mortalità	Prodotto tra l'incidenza della diagnosi e l'eccesso di rischio: $incidenza \cdot (CSMR_{diagnosticato} - CSMR_{non-diagnosticato})$ $= incidenza \cdot eccesso di rischio$ $= \frac{A+B}{A+B+C+D} \left(\frac{A}{A+B} - \frac{C}{C+D} \right)$ Nel quale: CSMR = tasso di mortalità causa-specifico																

Le analisi saranno condotte per tutte le 288 cause e condizioni stimate dal GBD e stratificate per età, sesso, anno, e ove disponibili, nei dati amministrativi regionali, livello di istruzione e reddito.

NOTA: La qualità dei dati amministrativi regionali è un prerequisito per la qualità delle stime prodotte. L'assessment della completezza e della copertura delle fonti dati verrà effettuato nella fase iniziale del progetto per ciascuna Regione partecipante.

5. Esiti attesi e significato

Il progetto si propone di produrre i seguenti risultati:

1. Stime della sopravvivenza per ogni malattia nell'arco di oltre 20 anni, utilizzando metodi diversi per età, sesso, luogo, anno, livello di istruzione e livello di reddito. Con questo lavoro e fornendo uno strumento di visualizzazione completo che, da una prospettiva basata sui dati, rivela il rischio relativo di morte e l'eccesso di mortalità associato a qualsiasi coppia diagnosi-causa di morte, sarà possibile avere una visione più olistica dei fattori che determinano gli esiti di salute in Italia e aiutare il sistema sanitario a pianificare in modo più efficace le politiche sanitarie del Paese.
2. Grazie all'approccio uniforme di mappatura delle cause e di pulizia dei dati adottato, i risultati ottenuti dalle Regioni italiane potranno essere confrontati con quelli di Norvegia, Brasile, Svezia, Danimarca e Colombia e potranno essere utilizzati per il miglioramento dei modelli e delle stime GBD.

3. Questo progetto migliorerà la condivisione delle conoscenze sulla metodologia adottata nel Global Burden of Disease Study (ad esempio, per la redistribuzione dei garbage codes).
4. Pubblicazione di articoli scientifici derivanti da queste analisi che esplorano la sopravvivenza comparativa e la perdita di capitale umano tra popolazioni diagnosticate e non diagnosticate per specifiche condizioni.
5. Rafforzare il ruolo dell'Italia come Paese attivo e riconosciuto nei network internazionali di analisi del carico di malattia e dei sistemi sanitari, promuovendo l'uso dei dati amministrativi per la valutazione comparativa delle politiche sanitarie e sostenendo processi di apprendimento reciproco tra Paesi.

6. Sostenibilità

Una volta consolidata l'infrastruttura e il coordinamento interregionale su una piattaforma condivisa, il sistema presenta le condizioni per una sostenibilità nel tempo, favorendone il mantenimento e l'ulteriore sviluppo.

7. Aspetti etici e protezione dei dati

Il progetto utilizza dati amministrativi sanitari già raccolti per finalità istituzionali, trattati nel rispetto della normativa vigente in materia di protezione dei dati personali (GDPR e normativa nazionale). Gli obiettivi di sanità pubblica dello studio giustificano l'uso dei dati, finalizzato a migliorare la conoscenza del carico di malattia e a supportare la programmazione sanitaria.

I dati saranno gestiti e analizzati esclusivamente in loco, presso le Regioni e le istituzioni competenti, senza trasferimento di dati individuali a server esterni. Le analisi saranno condotte su dati pseudonimizzati e i risultati diffusi solo in forma aggregata, garantendo la tutela della riservatezza degli individui.

In particolare, le seguenti garanzie tecniche e organizzative sono adottate:

- Nessun trasferimento di dati individuali: i dati individuali non vengono mai trasferiti ai server IHME o a server esterni. La pipeline analitica viene eseguita esclusivamente in loco, sui server regionali.
- Pseudonimizzazione: tutti i dati sono trattati mediante identificatori univoci pseudonimizzati. Non sono trattati dati direttamente identificativi in chiaro.
- Analisi in loco: il codice analitico sviluppato dall'IHME viene trasferito alle Regioni ed eseguito direttamente sulle loro infrastrutture locali.
- Risultati aggregati: solo risultati aggregati e anonimi vengono condivisi con il team IHME e utilizzati per le pubblicazioni scientifiche.
- Conformità GDPR: piena aderenza al GDPR (Regolamento (UE) 2016/679) e alla normativa italiana vigente in materia di protezione dei dati personali e ricerca scientifica.

Nei Paesi in cui le normative nazionali prevedono una revisione dei requisiti di condotta etica (ad esempio, da parte dei comitati etici), il NSC/NST deve soddisfare tali requisiti. Per il presente progetto è previsto il parere del DPO dell'ISS e la successiva approvazione del Comitato Etico Nazionale dell'ISS, il cui parere positivo copre tutte le Regioni partecipanti.

Ogni Regione partecipante è tenuta a:

- rispettare le normative regionali e nazionali applicabili in materia di protezione dei dati;
- garantire che i dati amministrativi siano trattati esclusivamente per le finalità del presente protocollo;
- documentare completamente le proprie procedure di accesso ai dati e di pseudonimizzazione;
- identificare un referente regionale per la protezione dei dati.

8. Team di ricerca e coordinamento

	Istituzione	Ruolo
Dr.ssa Benedetta Armocida	ISS, Dipartimento MACA	Responsabile scientifico per l'ISS
Dr. Marco Silano	ISS, Dipartimento MACA	Direttore dipartimento MACA
Dr.ssa Lisa Leonardini	ProMIS	Coordinatrice Comitato Tecnico Operativo ProMIS
Dr. Lorenzo Monasta	IRCCS Burlo Garofolo	Coordinatore Italian GBD Initiative
Dr. Mohsen Naghavi	IHME	Responsabile del team IHME
Dr.ssa Eve Wool	IHME	Responsabile della ricerca IHME

9. Regioni partecipanti

Le Regioni che hanno espresso interesse a partecipare al progetto nella fase iniziale sono le seguenti:

Regione	Note
Friuli-Venezia Giulia	Regione pilota – esperienza consolidata con sistemi informativi sanitari regionali integrati.
Marche	
Provincia Autonoma di Trento	Partecipazione in qualità di Provincia Autonoma; sistema informativo sanitario di riferimento.
Sicilia	

NOTA: La definizione dei referenti tecnici regionali e la mappatura delle fonti di dati disponibili per ciascuna Regione saranno oggetto della Fase 2 del progetto. Il progetto è aperto all'adesione di ulteriori Regioni nelle fasi successive.

10. Documenti e file da produrre

Al termine delle rispettive fasi operative, le Regioni partecipanti e il team di coordinamento produrranno i seguenti documenti e file:

- Dataset armonizzati e linkati per ciascuna Regione partecipante, conformi alle specifiche tecniche IHME.
- Stime di sopravvivenza, Rischio Relativo ed Eccesso di Mortalità per diagnosi, per tutte le 288 cause GBD.
- Report di mappatura delle fonti dati per Regione.
- Strumento di visualizzazione interattiva (dashboard) per il decision making in sanità pubblica.
- Report finale di progetto con confronto internazionale.
- Almeno un articolo scientifico sottomesso a una rivista ad alto impatto, con co-authorship regionale.
- Moduli di reporting compilati per il coordinamento nazionale e internazionale (ISS-MACA, IHME).

Si prega di contattare il NSC (ISS-MACA) per informazioni sulle specifiche tecniche dei file e sulle modalità di consegna.

Bibliografia

1. Souto Melo AP, Dippenaar I, Johnson SC, Weaver ND et al. *All-cause and cause-specific mortality among people with severe mental illness in Brazil's public health system, 2000-15: a retrospective study*. Lancet Psychiatry 2022.
2. GBD 2019 Diseases and Injuries Collaborators. *Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019*. Lancet 2020; 396: 1204–22.