

DATA SCIENCE FOR HEALTH

Studiare la rappresentatività degli studi clinici per sviluppare cure più efficaci per tutti

Bando con scadenza
30 gennaio 2024

BANDI 2023

Fondazione
CARIPLO

TUTE SERVARE MUNIFICE DONARE • 1816



Bando in collaborazione con



Indice

1. <i>PREMESSA</i>	4
2. <i>IL CONTESTO</i>	4
3. <i>GLI OBIETTIVI</i>	4
4. <i>LINEE GUIDA</i>	5
4.1 Soggetti destinatari.....	5
4.2 Progetti ammissibili.....	5
4.3 Criteri.....	7
4.4 Progetti non ammissibili.....	7
4.5 Iter di presentazione.....	8
5. <i>BUDGET DISPONIBILE, SPESE AMMISSIBILI E/O INAMMISSIBILI</i>	8
6. <i>COMUNICAZIONI E INFORMAZIONI</i>	8
7. <i>SINTESI</i>	8
8. <i>APPENDICE: STRUTTURA E INFORMAZIONI RELATIVE AL DATASET MESSO A DISPOSIZIONE DA NOVARTIS FARMA S.p.A.</i>	9

1. PREMESSA

“Data Science for Health - Studiare la rappresentatività degli studi clinici per sviluppare cure più efficaci per tutti” è un bando con scadenza emesso dall’Area Ricerca Scientifica e Trasferimento tecnologico che si inserisce nell’obiettivo strategico “Ricerca scientifica: sostenere ricerca multidisciplinare per il benessere delle persone e dello sviluppo socio-economico delle comunità”.

Il bando è promosso in collaborazione con Novartis Farma S.p.A. che rende accessibile il dataset dettagliato in Appendice per fini di pubblica utilità e di ricerca.

2. IL CONTESTO

La salute ed il benessere della popolazione sono beni inestimabili e la ricerca clinica costituisce uno strumento potente per promuoverli. Gli studi clinici, infatti, hanno lo scopo di ottenere prove oggettive in merito a efficacia e sicurezza di nuove terapie o procedure diagnostiche, con la finalità ultima di migliorare la salute del singolo e dell’intera comunità. Sebbene in linea teorica la partecipazione a tali studi sia aperta a tutti, questa è soggetta al rispetto di criteri di inclusione ed esclusione la cui progettazione e valutazione è assai complessa. Come recentemente precisato da uno studio pubblicato su Nature, criteri di ammissibilità eccessivamente restrittivi possono rappresentare una barriera alla validità dello studio stesso¹. In aggiunta, la U.S. Food and Drug Administration sottolinea che certe popolazioni sono solitamente escluse o sottorappresentate dagli studi clinici.

¹ Evaluating eligibility criteria of oncology trials using real-world data and AI. Nature, 592:629–633 (2021).

² Enhancing the Diversity of Clinical Trial Populations - Eligibility Criteria, Enrollment Practices, and Trial Designs Guidance for Industry (<https://www.fda.gov/media/127712/download>).

Inoltre, spesso si assiste ad un fenomeno di autoselezione dei pazienti in studio dovuto a ragioni culturali, sociali, logistiche o di disponibilità tempo. Ciò comporta che, talvolta, efficacia e sicurezza vengano testate su soggetti non rappresentativi dell’intera popolazione che potenzialmente potrebbe beneficiare di quel farmaco e che realmente lo utilizzerà dopo l’approvazione². Un’ulteriore recente pubblicazione evidenzia come inclusività e diversità siano sfide importanti per chi si occupa di ricerca clinica: la stratificazione aumenta la probabilità di rilevare effetti significativi, chiarisce parte della variabilità che si osserva, riduce i fattori confondenti e favorisce la riproducibilità³.

3. GLI OBIETTIVI

Alla luce del contesto precedentemente delineato, il bando intende sostenere progetti di ricerca nel campo della Scienza dei Dati che analizzino il gap di rappresentatività tra le popolazioni arruolate negli studi clinici e la popolazione generale affetta dalla medesima patologia.

A partire dal dataset reso disponibile da Novartis Farma S.p.A. e dettagliato in Appendice, i proponenti dovranno: esplicitare ipotesi conoscitive originali, rappresentare in maniera precisa le tecniche e i metodi della Scienza dei Dati che intendono applicare ed elaborare modelli e linee guida per sviluppare cure più efficaci per tutti. Nel caso in cui, nell’ambito della ricerca proposta, verranno utilizzati ulteriori dataset per analisi di confronto, i proponenti avranno cura di descriverli in dettaglio precisando il valore aggiunto per il raggiungimento degli obiettivi progettuali.

Qualora le proposte vengano presentate in partenariato, sarà essenziale chiarire le

³ Inclusion and diversity in clinical trials: Actionable steps to drive lasting change. Contemp Clin Trials, 116:106740 (2022).

competenze apportate da ciascun partner di progetto e come queste si integrano a quelle dell'ente capofila in un disegno unitario.

Al fine di stimolare il dibattito pubblico e promuovere l'adozione dei modelli e delle linee guida sviluppate, i proponenti dovranno individuare specifiche modalità di condivisione delle evidenze ottenute ad esito delle ricerche, produrre position paper e policy brief e coinvolgere gli stakeholder di riferimento in coerenza con i principi della Ricerca e Innovazione Responsabile (RRI)⁴.

Da ultimo, stante l'elevata richiesta da parte del mercato del lavoro di profili e competenze nel campo della Scienza dei Dati, le proposte dovranno offrire opportunità formative e di crescita professionale ai ricercatori coinvolti.

4. LINEE GUIDA

4.1 Soggetti destinatari

Enti pubblici o privati non profit che svolgono attività di ricerca scientifica. Unicamente in qualità di partner, potranno essere coinvolti anche Enti di Terzo Settore quali, ad esempio, le associazioni di pazienti. I soggetti capofila dovranno necessariamente possedere una struttura operativa situata sul territorio di prioritario riferimento della Fondazione, ovvero Lombardia e province di Novara e del Verbano Cusio Ossola. Eventuali partner di progetto potranno essere localizzati anche in altri territori.

Con la partecipazione al bando, gli enti consentono la portabilità del contributo per i responsabili scientifici. Il trasferimento della ricerca e dei fondi presso un altro ente saranno di volta in volta valutati dalla Fondazione, tenendo conto di quanto disposto nel testo del presente bando, nella "Guida alla presentazione dei progetti su bandi" e nei

⁴ Il movimento europeo che caratterizza il tema della RRI è attivo oramai da anni, soprattutto nel contesto europeo (https://ec.europa.eu/research/swafs/pdf/rome_declaration_RRI_final_21_November.pdf). Ad oggi, è possibile descrivere la RRI come un processo dinamico e iterativo che intende allineare la ricerca e l'innovazione ai valori, bisogni e aspettative dei cittadini. La RRI, inoltre, mira a coinvolgere attivamente tutti gli attori diversamente impegnati nelle pratiche della ricerca e innovazione, rendendoli mutualmente responsabili sia relativamente al processo di ricerca sia relativamente ai risultati prodotti.

⁵ Sono previste estensioni al periodo di eleggibilità:

"Criteri generali per la concessione dei contributi" disponibili sul sito web della Fondazione per il download.

4.2 Progetti ammissibili

Per essere considerati ammissibili alla valutazione, i progetti dovranno essere in possesso dei seguenti requisiti:

- sviluppare ricerca indipendente per fini di pubblica utilità;
- proporre un disegno di ricerca che utilizzi metodi e tecniche della Scienza dei Dati;
- utilizzare il dataset messo a disposizione da Novartis Farma S.p.A.;
- rispettare le priorità scientifiche così come dettagliate nel paragrafo 3 "Gli obiettivi";
- identificare come responsabile scientifico dell'unità capofila un ricercatore con un'età massima di 45 anni, compiuti entro la data di scadenza del bando⁵.

I responsabili di unità (capofila e partner):

- dovranno presentare sul presente bando una sola proposta progettuale;
- potranno presentare ulteriori richieste di contributo su altri strumenti dell'Area Ricerca come responsabili di unità solo nel caso in cui non vi sia sovrapposizione progettuale⁶;
- non dovranno risultare responsabili di unità in progetti ancora in corso⁷ finanziati nell'ambito di bandi dell'Area Ricerca (esclusi bandi in partenariato con altre agenzie di finanziamento).

I progetti che coinvolgeranno soggetti umani, materiale biologico umano, campioni umani e dati

- maternità: 18 mesi per ciascun figlio;
- paternità: effettivo periodo di congedo di paternità di cui si è usufruito;
- gravi malattie (oltre i 90 giorni): effettivo periodo di congedo per gravi malattie di cui si è usufruito.

⁶ In caso di sovrapposizione, sarà ritenuta ammissibile unicamente la prima proposta progettuale pervenuta.

⁷ Sono considerati in corso i progetti per i quali non sia ancora stata inoltrata la rendicontazione a saldo - scientifica ed economica - mediante caricamento sull'area riservata del sito web di Fondazione Cariplo.

personali di persone fisiche dovranno essere conformi alle leggi italiane in materia.

Il costo totale della ricerca proposta dovrà prevedere esclusivamente i costi addizionali di progetto⁸ e le spese correnti, nel rispetto dei criteri di eleggibilità e dei massimali di seguito elencati:

- A03 - “Acquisto di arredi e attrezzature”
Tale voce di spesa non dovrà superare la soglia del 30% dei costi addizionali di progetto e potrà riguardare il costo totale sostenuto per l’acquisto di attrezzature o programmi software di uso pluriennale acquisiti ex novo, di cui si dimostri la reale necessità per la realizzazione del piano sperimentale.
- A04 - “Altre spese per investimenti ammortizzabili”
Tale voce di spesa dovrà riguardare i costi di brevettazione e gli importi da sostenere nell’ipotesi di affitto di attrezzature.
- A06 - “Personale non strutturato”
Tale voce di spesa dovrà comprendere il solo personale addetto alla ricerca, con esclusione quindi di figure cui competano ruoli amministrativi.
- A07 - “Prestazioni professionali di terzi”⁹
Tale voce di spesa potrà riguardare i costi di accesso a basi di dati e servizi di utilità per il progetto quali, a titolo esemplificativo, potenza di calcolo o macchine virtuali in *cloud* e remoto.
- A08 - “Materiali di consumo”
Tale voce di spesa non dovrà comprendere cancelleria d’ufficio e fotocopie.

⁸ Per “costi addizionali di progetto” si intende la sommatoria delle voci A03, A04, A06, A07, A08, A10. In altre parole, il contributo richiesto meno la voce A09.

⁹ Rientrano in questa voce di spesa anche gli eventuali costi per la certificazione del professionista abilitato iscritto al Registro dei Revisori Legali. Ai soli fini del calcolo finalizzato ad accertare se sussiste o meno l’obbligo di acquisire la certificazione (audit), si precisa che l’importo da assumere a riferimento si ottiene moltiplicando i costi inseriti nel Piano Economico (sommatoria delle voci A06, A07, A08, A09 e A10) per il coefficiente 1,25. Si raccomanda, inoltre, di fare riferimento alla “Guida alla Rendicontazione” e alle “Linee guida per l’affidamento dell’incarico di audit”; entrambi i documenti sono consultabili sul sito internet. Ad ogni modo, si anticipa che in caso di

- A09 - “Spese correnti”
Tale voce di spesa non dovrà superare la soglia del 5% dei costi addizionali di progetto.
- A10 - “Altre spese gestionali”
Tale voce di spesa potrà comprendere i costi per missioni, partecipazioni a congressi, meeting tra partner, pubblicazioni, attività formative e disseminazione.

La Fondazione garantirà la copertura del 100% dei costi di progetto; pertanto, il contributo richiesto dovrà coincidere con il costo totale.

La richiesta di contributo non dovrà essere superiore a 200.000 Euro.

Per quanto concerne la tempistica, i progetti dovranno avere durata massima di 24 mesi.

Il bando richiede che la proposta sia corredata dei seguenti documenti obbligatori:

- Lettera accompagnatoria¹⁰;
- Accordo di partenariato¹¹;
- Descrizione dettagliata del progetto, in formato PDF¹²;
- Piano economico, in formato EXCEL¹³.

Con la partecipazione al bando, l’organizzazione accetta e si impegna a rispettare le indicazioni contenute nei seguenti documenti: “Policy della Fondazione Cariplo in tema di tutela della proprietà intellettuale”, “Policy di open access” e “Linee guida per la citazione del contributo nelle comunicazioni scientifiche” disponibili online per il download. Si rammenta che tutte le pubblicazioni che deriveranno dai risultati di progetto dovranno riportare l’affiliazione dell’organizzazione con cui la proposta progettuale è stata presentata. Inoltre, ai

finanziamento la Fondazione confermerà ai singoli beneficiari l’eventuale obbligo di soddisfare l’adempimento in questione.

¹⁰ La Lettera accompagnatoria dovrà essere redatta sulla base del form disponibile online per il download.

¹¹ L’Accordo di partenariato sarà un documento obbligatorio solo nel caso in cui siano presenti partner di progetto e dovrà essere redatto sulla base dei form disponibili on line per il download.

¹² La Descrizione dettagliata del progetto dovrà essere redatta sulla base del form messo a disposizione per il bando.

¹³ Il Piano economico in formato Excel verrà automaticamente generato dalla piattaforma all’interno della sezione “Piano economico”.

fini della predisposizione della proposta, si raccomanda un'attenta lettura delle FAQ appositamente redatte per il presente bando.

4.3 Criteri

La valutazione dei progetti pervenuti sarà svolta secondo le modalità indicate nella "Guida alla presentazione dei progetti su bandi"¹⁴.

L'esame di merito sarà condotto sulla base dei criteri di seguito rappresentati:

1. QUALITÀ SCIENTIFICA (peso 30%)

- chiarezza nella formulazione degli obiettivi e delle strategie adottate;
- adeguata contestualizzazione rispetto allo stato dell'arte;
- presenza di un disegno di ricerca logico e ben strutturato;
- adeguatezza dei metodi e delle tecniche proposte rispetto agli obiettivi conoscitivi;
- originalità del progetto in termini di metodi e tecniche proposte.

2. RICADUTE (peso 25%)

- impatto atteso in termini di avanzamento della conoscenza;
- impatto atteso in termini di benefici per i cittadini e la collettività;
- previsione di adeguate modalità di condivisione dei risultati della ricerca e capacità di alimentare un dibattito pubblico sul tema coinvolgendo gli stakeholder di riferimento;
- impatto atteso in termini di contributo per l'elaborazione di nuovi modelli e linee guida.

3. RESPONSABILE SCIENTIFICO, GRUPPO DI LAVORO E ORGANIZZAZIONE (peso 10%)

- leadership scientifica e gestionale del responsabile scientifico dell'unità capofila;
- adeguate competenze dei membri del team di ricerca;

- ricadute del progetto in termini di crescita scientifica e professionale;
- dotazione di infrastrutture e strumenti adeguati alla dimensione e tipologia del progetto.

4. RILEVANZA PER LA FONDAZIONE (peso 15%)

- impatto atteso in termini di creazione di valore per la comunità locale.

5. ADEGUATEZZA RISPETTO AL DATASET (peso 10%)

- compatibilità dell'ipotesi di ricerca e degli obiettivi progettuali rispetto alle caratteristiche del dataset messo a disposizione.

6. PIANO ECONOMICO E DURATA (peso 10%)

- adeguatezza e coerenza del Piano economico e della durata del progetto.

Si precisa che la valutazione di merito dei progetti sarà effettuata mediante *peer review*, ovvero attraverso una valutazione scientifica di merito demandata a una parte terza e indipendente costituita da esperti qualificati che sottostanno a regole e procedure precisamente definite e volte a escludere conflitti di interesse, pregiudizi ideologici, pressioni personali e autoreferenzialità. Gli Uffici di Fondazione Cariplo esprimeranno un giudizio limitatamente alla rilevanza per la Fondazione e alla coerenza del Piano economico. Novartis Farma S.p.A. valuterà l'adeguatezza rispetto al dataset.

4.4 Progetti non ammissibili

Saranno ritenuti inammissibili alla valutazione i progetti che, oltre a non rispettare quanto indicato al precedente paragrafo 4.2 "Progetti ammissibili", presentino le seguenti caratteristiche:

- ricerche che non contemplino la generazione di nuova conoscenza a vantaggio del bene comune;
- progetti unicamente finalizzati allo sviluppo di dataset, software, nuovi modelli matematici e strumenti predittivi di analisi senza riferimento

¹⁴ A partire dal 2020, la Fondazione intende gradualmente incrementare l'attenzione verso gli impatti ambientali e climatici

dei progetti finanziati. Per maggiori approfondimenti si veda il paragrafo 10 della suddetta guida.

alle priorità scientifiche del bando, così come dettagliate nel paragrafo 3 “Gli obiettivi”;

- richieste di finanziamento per la creazione di nuovi centri di ricerca.

4.5 Iter di presentazione

La domanda di contributo dovrà essere presentata entro e non oltre le ore 17 del **30 gennaio 2024**.

5. BUDGET DISPONIBILE, SPESE AMMISSIBILI E/O INAMMISSIBILI

Il budget a disposizione è pari a 800.000 Euro.

Con riferimento alle spese ammissibili e/o inammissibili, si rimanda a quanto indicato al precedente paragrafo 4.2 “Progetti ammissibili”.

6. COMUNICAZIONI E INFORMAZIONI

La Fondazione, in quanto soggetto privato, non è tenuta all’osservanza di procedure di evidenza pubblica e ha la facoltà di non assegnare in tutto o in parte il budget previsto. Il testo del bando e i documenti citati sono disponibili sul sito web di Fondazione Cariplo (www.fondazionecariplo.it).

Ai fini della valutazione di merito mediante *peer review*, i dati personali eventualmente riportati nella Descrizione dettagliata del progetto potranno essere trasferiti in Paesi o organizzazioni extra Ue, in conformità al Regolamento EU 2016/679. Prima di presentare domanda, dunque, si invita a leggere l’informativa privacy per avere maggiori informazioni sul trattamento dei dati personali e sul trasferimento dei dati personali verso Paesi o organizzazioni fuori dall’Unione europea.

7. SINTESI*

Bando	Data Science for Health
Tipo	Con scadenza
Scadenza	30 gennaio 2024
Budget disponibile	€ 800.000
Obiettivi	Il bando intende sostenere progetti di ricerca che applicano le tecniche e i metodi della Scienza dei Dati per analizzare la rappresentatività delle popolazioni arruolate negli studi clinici con la finalità ultima di sviluppare cure più sicure ed efficaci per tutti
Destinatari	Enti pubblici o privati non profit
Principali limiti di finanziamento	Contributo massimo € 200.000
Riferimenti	Area Ricerca Scientifica Contatti staff disponibili sul sito web www.fondazionecariplo.it

** I dati riportati al paragrafo “Sintesi” hanno mera valenza riassuntiva delle condizioni e clausole riportate nel testo del bando, alle quali si rinvia per una descrizione integrale.*

8. APPENDICE: STRUTTURA E INFORMAZIONI RELATIVE AL DATASET MESSO A DISPOSIZIONE DA NOVARTIS FARMA S.p.A.

1. Introduzione

Con l'obiettivo di costruire uno strumento di ricerca da condividere per fini di pubblica utilità, Novartis Farma S.p.A. ha estratto dati relativi ai pazienti arruolati in una serie di trial clinici coordinati dal dipartimento medico di Novartis e relativi a svariate patologie, nel completo rispetto della data privacy.

Dagli archivi del dipartimento medico di Novartis sono stati selezionati 26 studi di fase IIIb e fase IV, di cui 19 interventistici e 7 osservazionali, dai quali sono state estratte informazioni relative a circa 15000 pazienti e raccolte in un unico dataset in forma anonima. In particolare, si tratta di:

- studi conclusi;
- studi con numerosità pazienti ≥ 100 ;
- studi con *First Patient First Visit* (FPFV) consecutiva (\geq) al 01/01/2000.

2. Variabili considerate

Caratteristiche demografiche

AGE: età al momento della visita alla baseline. Nel caso in cui la variabile età non fosse presente, è stata derivata come differenza in anni tra la data di nascita e la data corrispondente alla visita alla baseline del relativo paziente.

SEX: sesso del paziente (M, F).

RACE: etnia del paziente (Asian, Black, Caucasian, Native American, Other, Pacific Islander, Unknown).

REGION: regione di appartenenza del centro in cui il paziente partecipa allo studio. Questo dato è stato ricavato a partire dalla provincia del centro di appartenenza con lo scopo di fornire un'indicazione geografica relativa al paziente. Nel rispetto della normativa sulla privacy, non sono disponibili dati relativi alla residenza dei pazienti.

DATE: mese e anno corrispondenti alla visita al basale.

Segni vitali

DBP (Dyastolic Blood Pressure): pressione diastolica misurata durante la visita alla baseline. Unità di misura: mmHg.

SBP (Systolic Blood Pressure): pressione sistolica misurata durante la visita alla baseline. Unità di misura: mmHg.

HR (Heart Rate): numero di battiti al minuto misurati durante la visita alla baseline. Unità di misura: battiti/min.

BMI (Body Mass Index): indice di massa corporea. In caso di non disponibilità di questa variabile, il suo valore è stato derivato dal rapporto tra il peso del paziente espresso in chilogrammi e il quadrato dell'altezza espressa in metri. Unità di misura: Kg/m^2 .

Analisi di laboratorio

VAR: codice/abbreviazione corrispondente alla variabile di laboratorio (es.: "ALB" = Albumin).

LABEL: nome in forma estesa della variabile di laboratorio.

LABVALUE: valore numerico corrispondente alla valutazione di laboratorio.

UNIT: unità di misura.

Ogni paziente all'interno del database presenterà una serie di valori di laboratorio riportati attraverso record multipli, in base ai dati di laboratorio disponibili. Pertanto, ad ogni paziente corrispondono più record, ognuno

dei quali contiene una valutazione di laboratorio. Il numero di record corrispondente ad un singolo paziente è quindi variabile in base alla quantità di valutazioni di laboratorio che sono state effettuate alla baseline.

Il database complessivo, così strutturato, presenta in totale 144623 record relativi a 15620 pazienti.

Le unità di misura corrispondenti ai dati di laboratorio sono state riportate all'interno della variabile "UNIT" (vedi tabella sottostante).

VAR1	Label	Evaluation	Unit
ALB	Albumin	Baseline	%, g/L, g/dL, mg/dL
ALP	ALP	Baseline	U/L
ALT	ALT	Baseline	U/L
AST	AST	Baseline	U/L
BAS	Basophils	Baseline	%, 10 ³ /mL, 10 ³ /μL, 10 ⁹ /L
CA	Calcium	Baseline	mEq/L, mg/dL, mmol/L
CHOL	Total cholesterol	Baseline	mg/dL, mmol/L
CL	Chlorine	Baseline	mEq/L, mmol/L
CREA	Serum Creatinine	Baseline	mg/dL, μmol/L
DBIL	Direct bilirubin	Baseline	mg/dL
EOS	Eosinophils	Baseline	%, 10 ³ /mL, 10 ³ /μL, 10 ⁹ /L
GLY	Glycemia	Baseline	mg/dL, mmol/L
HBA1C	Glycated hemoglobin	Baseline	%
HCT	Hematocrit	Baseline	%
HDL	HDL	Baseline	mg/dL, mmol/L
HGB	Hemoglobin	Baseline	g/dL, g/L
K	Potassium	Baseline	mEq/L, mmol/L
LDL	LDL	Baseline	mg/dL, mmol/L
LYM	Lymphocytes	Baseline	%, 10 ³ /mL, 10 ³ /μL, 10 ⁹ /L
MON	Monocytes	Baseline	%, 10 ³ /mL, 10 ³ /μL, 10 ⁹ /L
N	Nitrogen	Baseline	mg/dL
NEU	Neutrophils	Baseline	%, 10 ³ /mL, 10 ³ /μL, 10 ⁹ /L
OTH	Other	Baseline	%, 10 ⁹ /L
PLAT	Platelets	Baseline	10 ³ /μL, 10 ⁹ /L
RBC	Erythrocytes	Baseline	10 ¹² /L, 10 ³ /μL, 10 ⁶ /L, 10 ⁶ /μL
SOD	Sodium	Baseline	mmol/L, mEq/L
TBIL	Total Bilirubin	Baseline	mg/dL, μmol/L
TPROT	Total protein	Baseline	g/L, g/dL
TRI	Triglycerides	Baseline	mg/dL, mmol/L
UACID	Urate	Baseline	mg/dL, mmol/L, μmol/L
UREA	Urea	Baseline	mg/dL, g/L
WBC	Leukocytes	Baseline	10 ³ /μL, 10 ⁹ /L

Range di laboratorio

I valori di riferimento delle valutazioni di laboratorio possono cambiare in base ad età e sesso del paziente e alla tipologia di strumentazione in uso nel singolo centro. Pertanto, si potranno notare valori variabili sia tra centri diversi sia all'interno dello stesso centro.

LORNG: limite inferiore dell'intervallo di riferimento. Un valore inferiore a questo limite può essere definito come alterato.

HIRNG: limite superiore dell'intervallo di riferimento. Un valore superiore a questo limite può essere definito come alterato.

Variabili aggiunte

SUBJID: identificativo del paziente. Numero casuale attribuito ad ogni paziente, da 1 a 16557.

NB: il numero totale di pazienti è 15620 e non corrisponde con l'identificativo.

PATCOD: codice identificativo della patologia principale oggetto dello studio da cui i dati di quel paziente sono stati estratti.

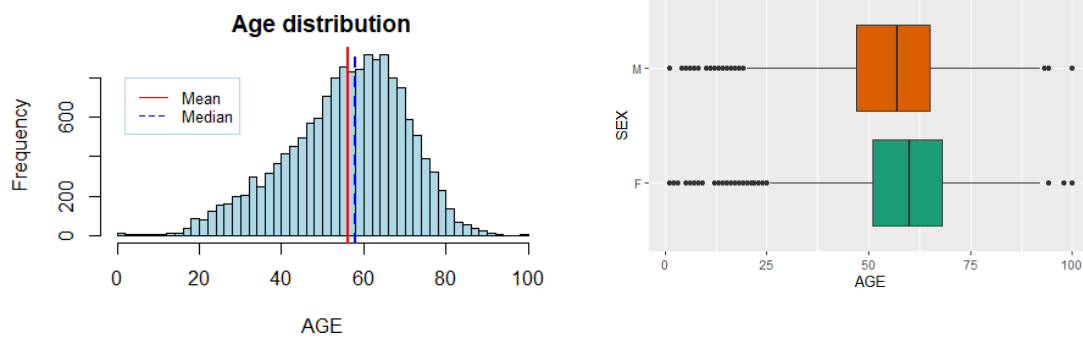
3. Panoramica del database

Il database contiene 144623 record e 17 colonne ed è in formato “long” (record multipli per ogni paziente) per la variabile “VAR” (variabile categoriale identificativa dell’esame di laboratorio svolto) e contiene dati su un campione totale di 15620 pazienti.

Pazienti

I pazienti hanno complessivamente un’età media di circa 56 anni e sono per il 55% maschi e 45% femmine. L’etnia dei pazienti è quasi interamente costituita da soggetti di razza caucasica: la proporzione di pazienti di etnia asiatica, nera, nativa americana e isolani del pacifico sono al di sotto dell’1%.

Stratificando per la variabile SEX, si ottiene un’età media per il sesso maschile (55,8) inferiore rispetto a quella del sesso femminile (56,7).



I pazienti che hanno partecipato agli studi selezionati sono stati arruolati principalmente in centri localizzati nelle regioni Campania, Lombardia e Sicilia (rispettivamente 1852, 1729 e 1449 pazienti). Il grafico a mappa rappresenta la distribuzione dei pazienti in base alla regione del centro di appartenenza.



Studi

La variabile PATCOD (codice patologia) rappresenta la patologia principale oggetto dello studio selezionato. La tabella riportata in seguito illustra la distribuzione di pazienti in base alla relativa patologia.

CNV	207	MetS	268
DIA_2	8918	PARTIAL SEIZURE	284
HBP	579	PSO_00	653
HEA_TRANS	385	PSO_01	431
KID_TRANS	2063	PSO_02	258
LIV_TRANS	186	SD	176
MS	1212		

Per ogni patologia, sono disponibili i criteri di inclusione ed esclusione agli studi.

Limiti

Le principali limitazioni del dataset sono attribuibili alla varietà delle caratteristiche degli studi considerati:

- l'inclusione degli studi osservazionali ha permesso di incrementare il numero di osservazioni ma ha generato un numero consistente di *missing values*;
- le valutazioni di laboratorio riportate derivano da numerosi centri italiani e sono riportate con unità di misura differenti.