

Database di laboratorio e outcome clinico

Presente e futuro della medicina di laboratorio

Bologna 6 dicembre 2019

Dott. Tommaso Trenti

Direttore dipartimento interaziendale attività integrata di medicina di laboratorio e anatomia patologica Azienda Ospedaliera
Universitaria e Azienda USL di Modena e membro del Direttivo Sibioc

Ing Debora Angeletti

Data & BI Chief Manager Azienda USL di Modena e membro del Direttivo AISIS

Dott. Tommaso Pirotti

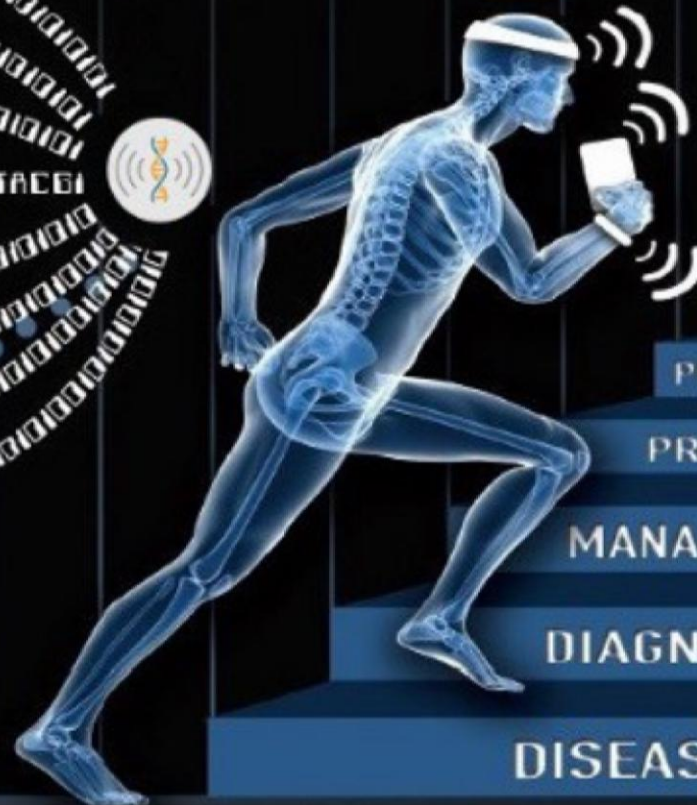
Dipartimento di Economia Marco Biagi, Università degli Studi di Modena e Reggio Emilia

THE DIGITAL HEALTH REVOLUTION

Infographic by Paul Sonnier

DIGITAL
REVOLUTION
+
GENOMIC
REVOLUTION

WIRELESS SENSORS & DEVICES
MOBILE CONNECTIVITY
SOCIAL NETWORKING
GENOMICS
INTERNET
IMAGING
DATA UNIVERSE
HEALTH INFO SYSTEMS



R
E
S
U
L
T
S

PREVENTION

PREDICTION

MANAGEMENT

DIAGNOSIS

DISEASE

1970 1980 1990 2000

storyofdigitalhealth.com

2010

2020



I **dati biometrici**, l'**alimentazione** e lo **stile di vita** assumono un ruolo decisivo per migliorare l'efficacia delle terapie, coinvolgendo anche l'**aspetto sociale** del paziente.

La **Medicina della persona** (o Medicina personalizzata) rappresenta il sistema di cura nell'era digitale ed è anche una possibile soluzione alle nuove problematiche del prossimo futuro, come l'**aumento dell'aspettativa di vita**, la **multi-morbilità**, la **cronicità delle malattie** ed il **sostenimento dei costi dell'intero sistema sanitario**

Il dato di laboratorio ha un ruolo di primaria importanza nei progetti di Digital Health in quanto offre un mare per lo più inesplorato di dati che possono essere analizzati e finalizzati alla conoscenza della popolazione e alla personalizzazione dei servizi.

Il principale produttore di dati strutturati aziendale

L'informazione su cui sono prese la maggior parte delle decisioni cliniche quindi centro nevralgico di controllo dell'appropriatezza clinica

Uno dei principali centri di costo aziendali, generatore di costi indotti, quindi centro nevralgico di controllo dell'appropriatezza prescrittiva

La principale interfaccia dell'Azienda verso la medicina del territorio e i cittadini

Il luogo dove sono applicate le più avanzate tecnologie diagnostiche e di data mining

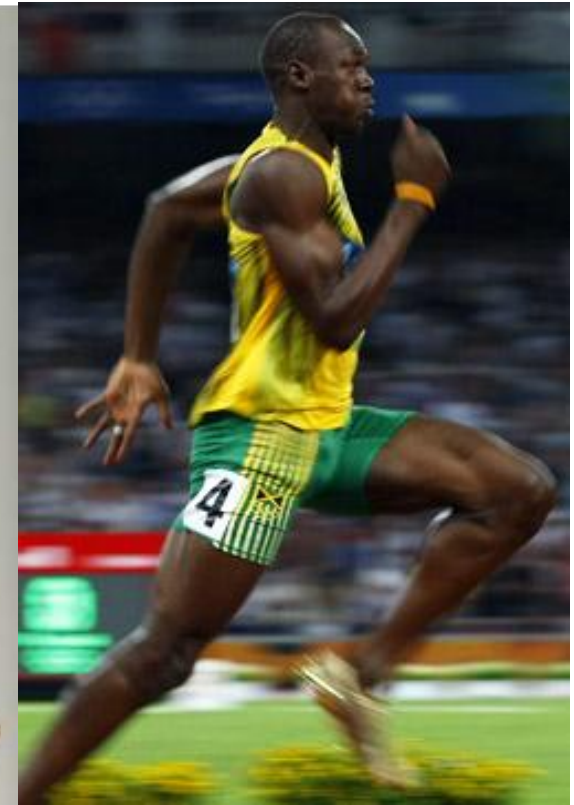
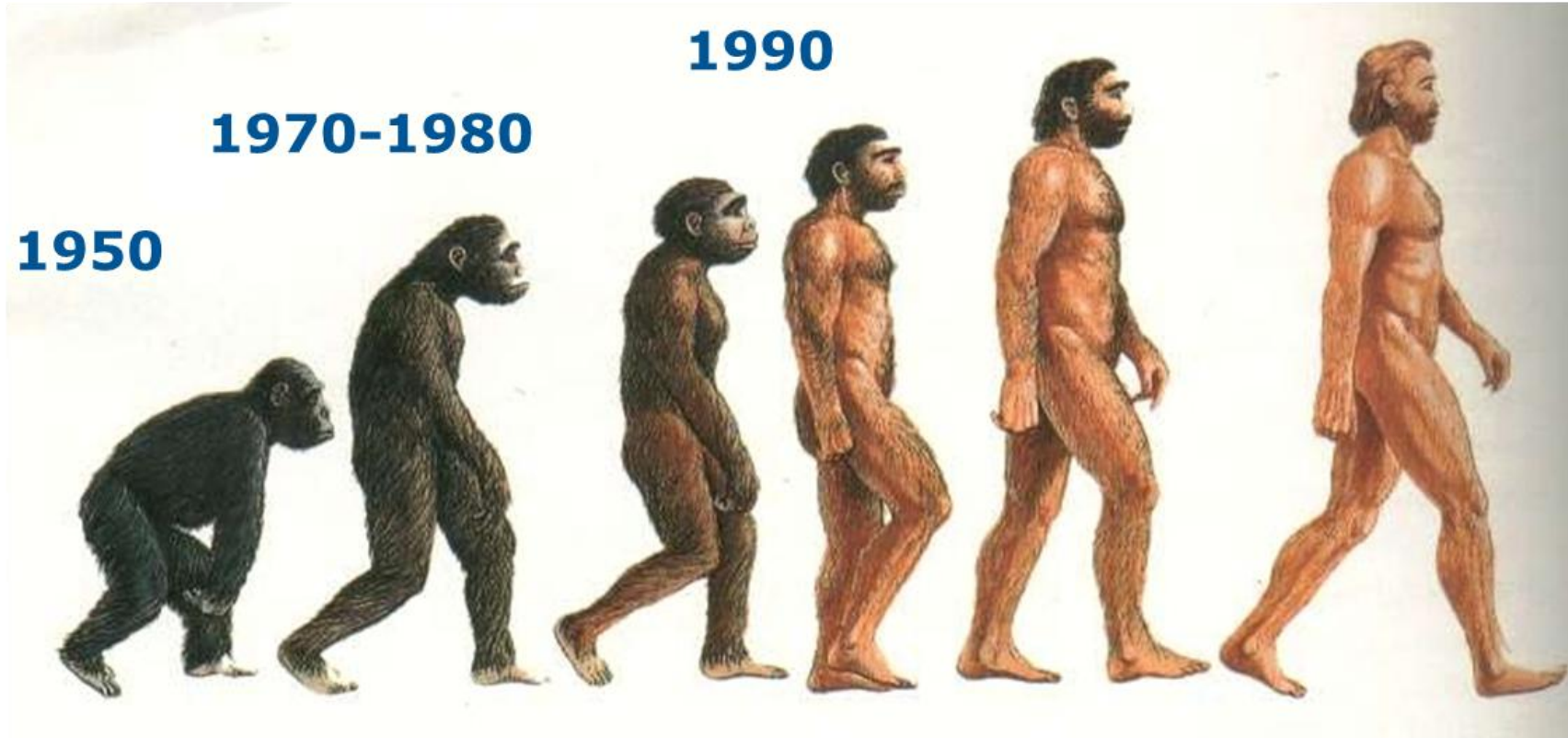
2020

2011

1990

1970-1980

< 1950



Indagini manuali

Strumenti Automatici

Workstation Integrate

Total Lab Automation

**AI Big Data
Data mining**

Dipartimento di Patologia Clinica



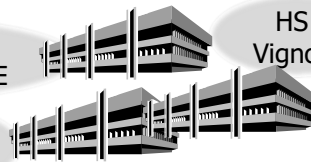
BLU
Laboratorio
Centrale
Provinciale

Esami per Esterni
della Provincia



Rete PoCT per
l'Emergenza

HS
Castelfranco E



HS
Vignola

HS
Sassuolo



Amb' Amb' Ambulatori

3 Laboratori tradizionali
Esami HOT-per ospedale

Laboratorio
Patologia Clinica
Ospedale di Pavullo



Laboratorio
Patologia Clinica
Ospedale di Mirandola



Laboratorio
Patologia Clinica
Ospedale di Carpi



Progetto: studio di appropriatezza dei valori di normalità

La definizione di “**normale**” per un esame di laboratorio impatta significativamente sul complessivo **processo diagnostico e terapeutico**.

In modo ideale i risultati di un esame di laboratorio dovrebbero essere interpretati avendo come riferimento una **popolazione di individui “simili” e “sani”**.

La disponibilità di **grandi quantità di dati** provenienti da fonti eterogenee e di tecniche di **machine learning** consente di ottenere notizie rappresentative della **condizione di salute**, non solo dei singoli come “**pazienti**” ma in generale delle “**comunità di cittadini**”



Quali sono i criteri per definire un **valore normale**?

Come identificare le molteplici e numerose peculiarità degli individui che compongono una **popolazione sana**?

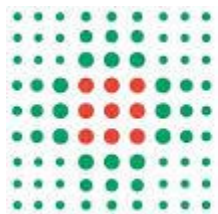
Quali e quanti sottoinsiemi è utile o necessario considerare nella classificazione della popolazione perché lo studio possa essere tradotto in **medicina di precisione**?

Qual è la **metodologia** più affidabile e adatta per rispondere a questi quesiti?

Di quali **competenze** è necessario avvalersi?

competenze scientifiche capaci di valutare e confrontare i risultati ottenuti nell'ambito del processo clinico diagnostico per evitare interpretazioni e conclusioni errate

competenze informatiche e statistiche capaci di valutare l'utilizzo proprio di metodi e tecnologie per garantire la robustezza matematica e tecnologica degli algoritmi applicati ed evitare che producano risultati affetti da errori



SERVIZIO SANITARIO REGIONALE
EMILIA-ROMAGNA
Azienda Unità Sanitaria Locale di Modena

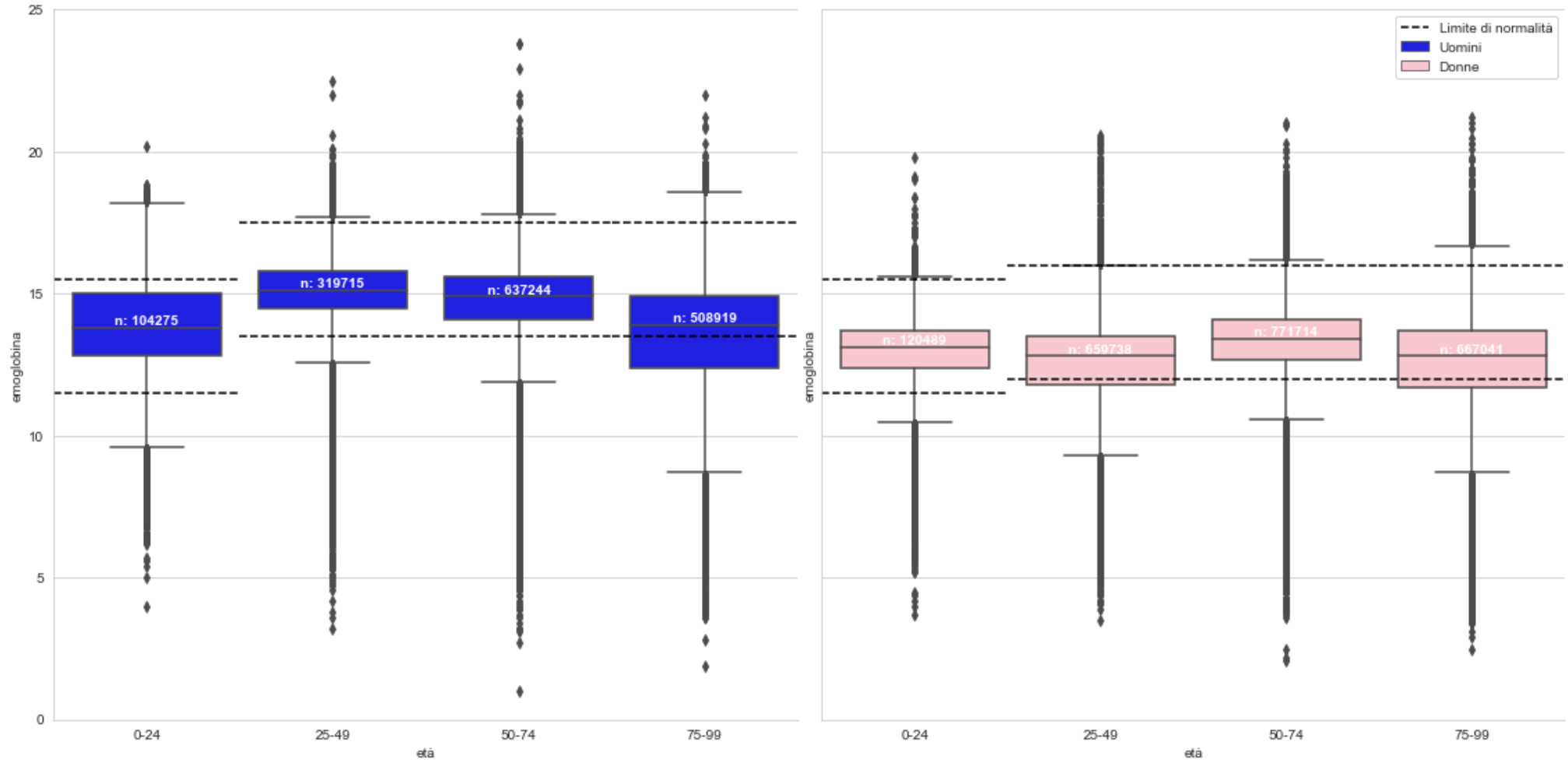


UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Fase 1

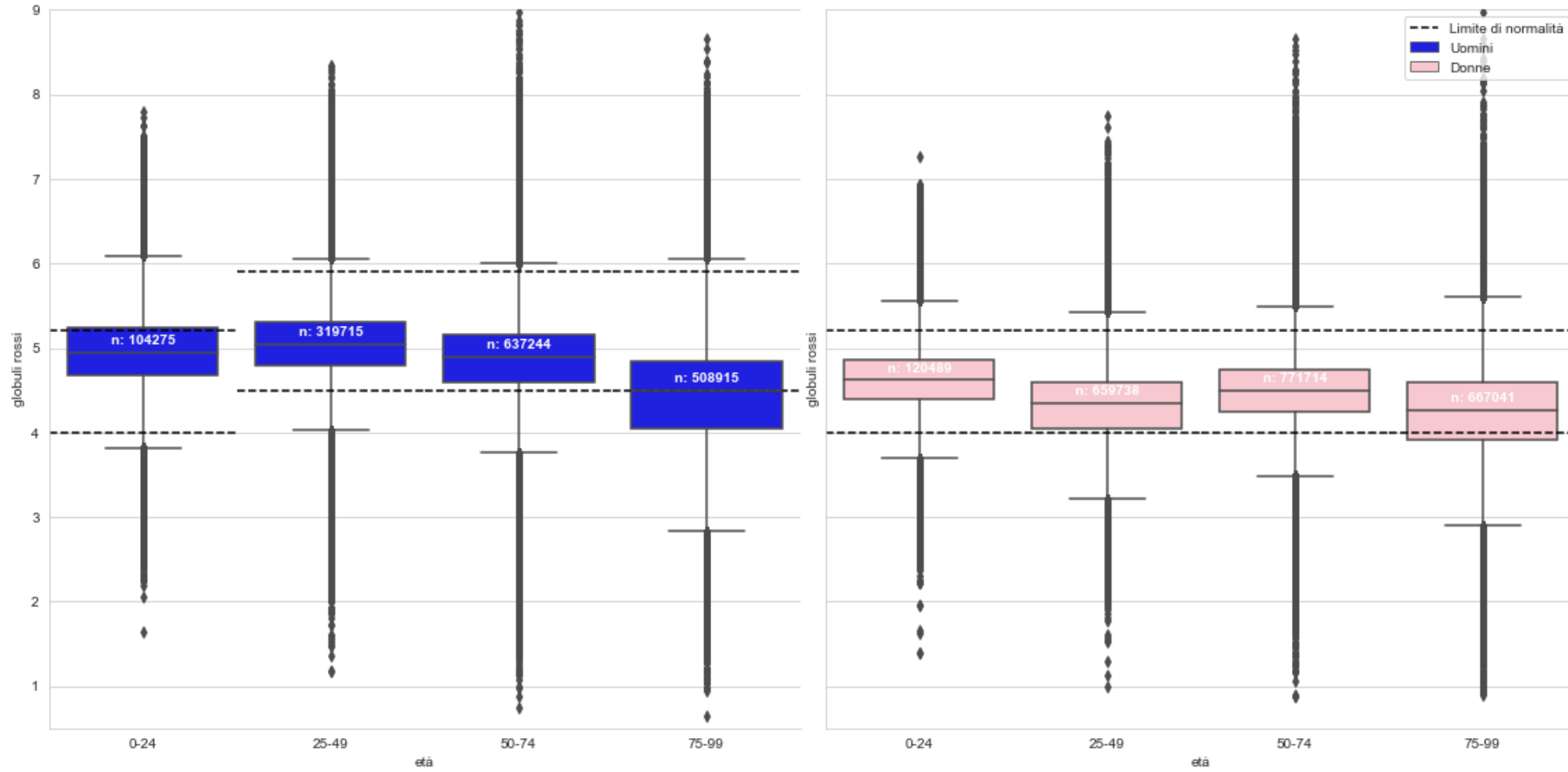
- Identificazione delle analisi oggetto di studio del comportamento rispetto ai valori soglia: emoglobina, numero di globuli rossi, di globuli bianchi, di piastrine e volume medio globulare
- A partire dal datawarehouse aziendale composizione del datamart di tutte le analisi effettuate dal 2010 ad oggi nei laboratori dell'AUSL di Modena non collegate ad episodi di ricovero: oltre **tre milioni** di singole osservazioni per ciascuna delle cinque analisi.
- Classificazione della popolazione relativa alle osservazioni per fasce di età e per genere e confronto delle osservazioni con le soglie di normalità.
- **Questo controllo ha evidenziato per globuli rossi e emoglobina che una porzione rilevante della popolazione è risultata al di fuori delle soglie di normalità e formalmente anemica.**

emoglobina: popolazione completa
osservazioni totali: 3791500



globuli rossi: popolazione completa

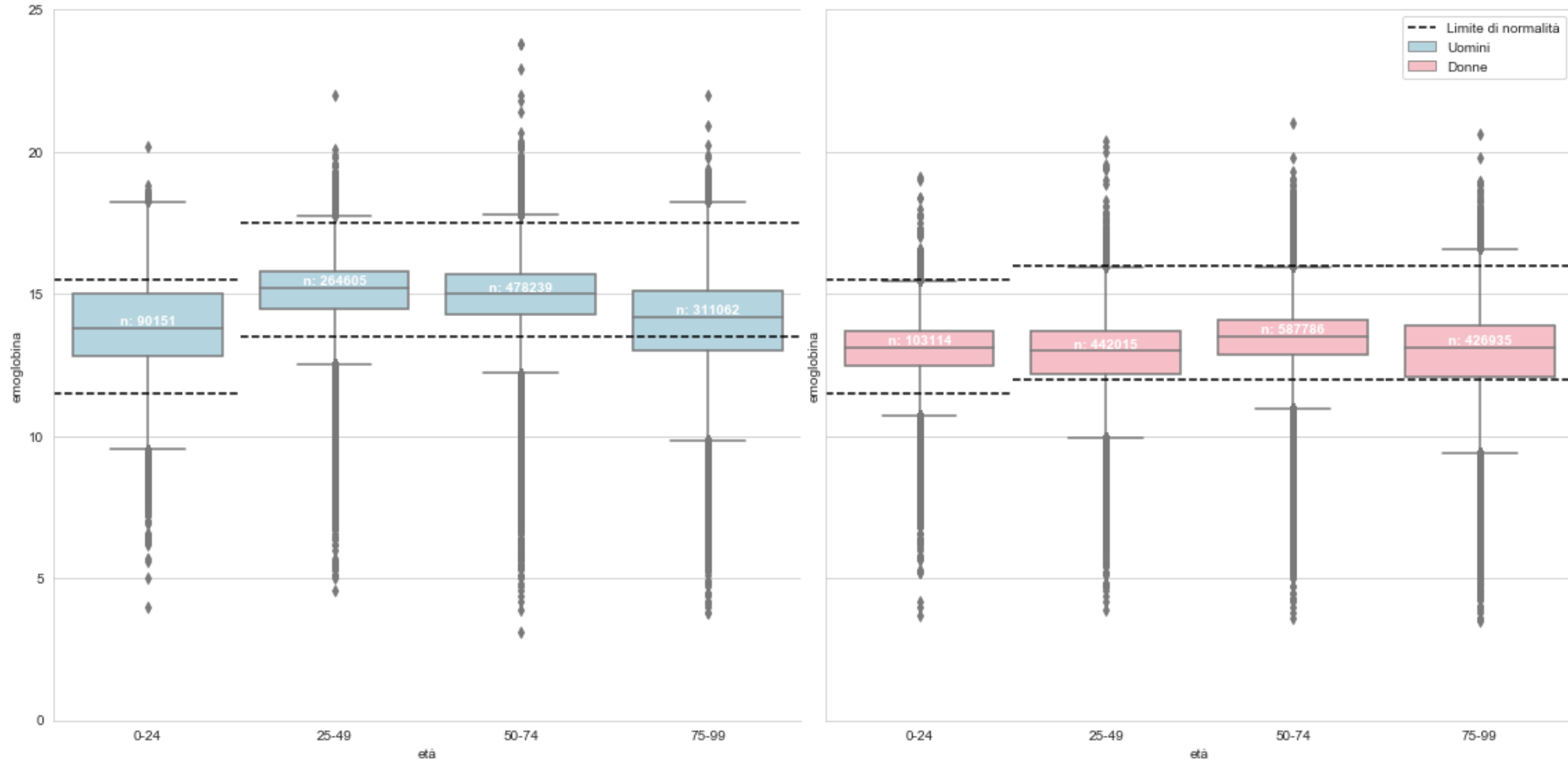
osservazioni totali: 3791496



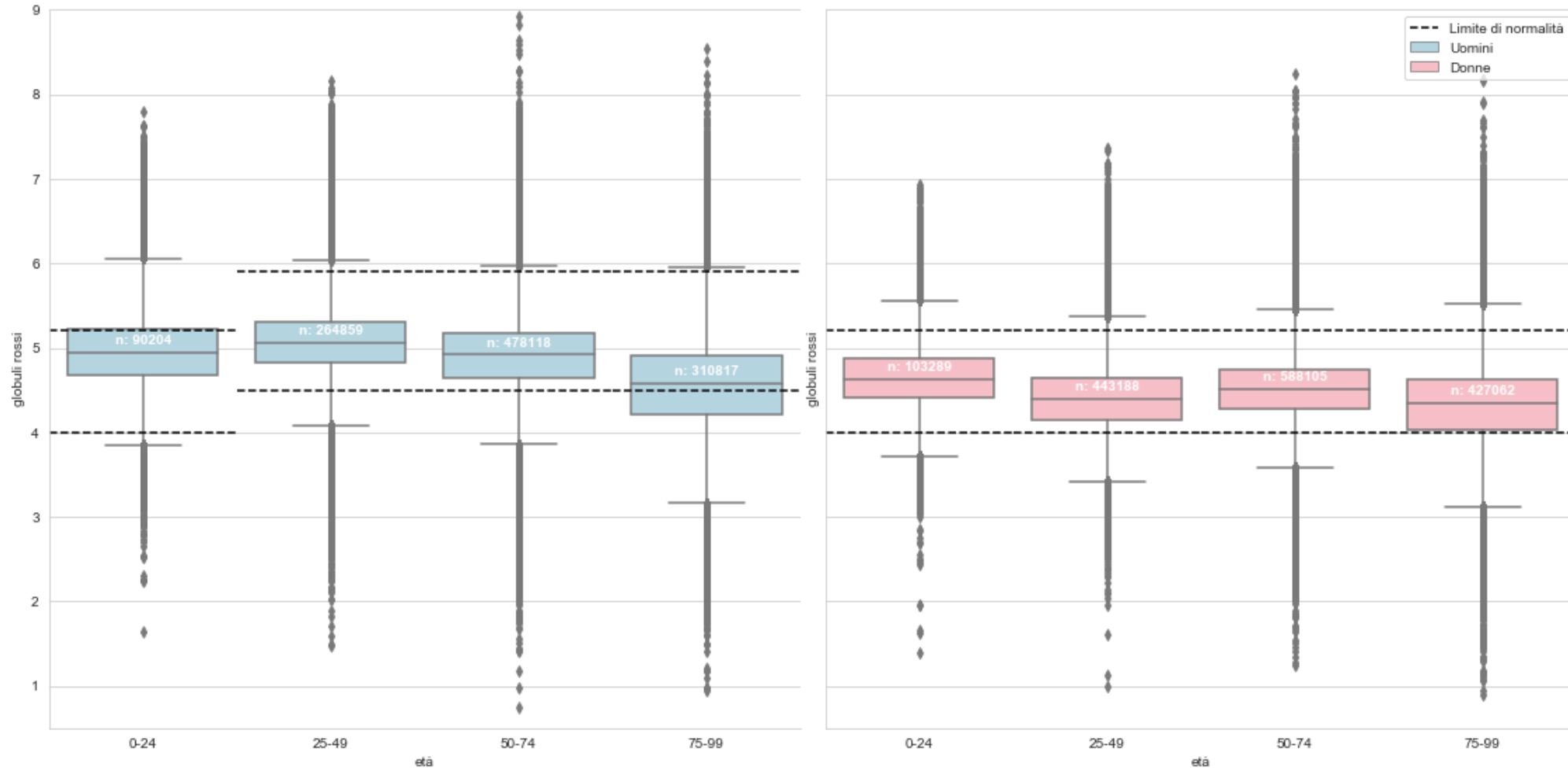
Fase 2: riduzione dei bias

- Per ridurre i possibili bias derivanti da analisi ripetute si è replicata la rilevazione utilizzando la **media annua delle analisi di ogni individuo** per ciascuna delle cinque caratteristiche ematiche in esame. Questo procedimento ha abbassato il numero di osservazioni a circa 2 milioni, ma ha sostanzialmente confermato il risultato ottenuto.
- **La tesi che emerge è che le soglie di normalità per globuli rossi ed emoglobina per le classi di età più anziane, come già avvenuto per la popolazione in età pediatrica, richiederebbero un sostanziale riesame perché non sono in grado di rilevare il fisiologico calo di globuli rossi ed emoglobina della popolazione anziana.**

emoglobina: popolazione completa (valori annui medi)
osservazioni totali: 2704685



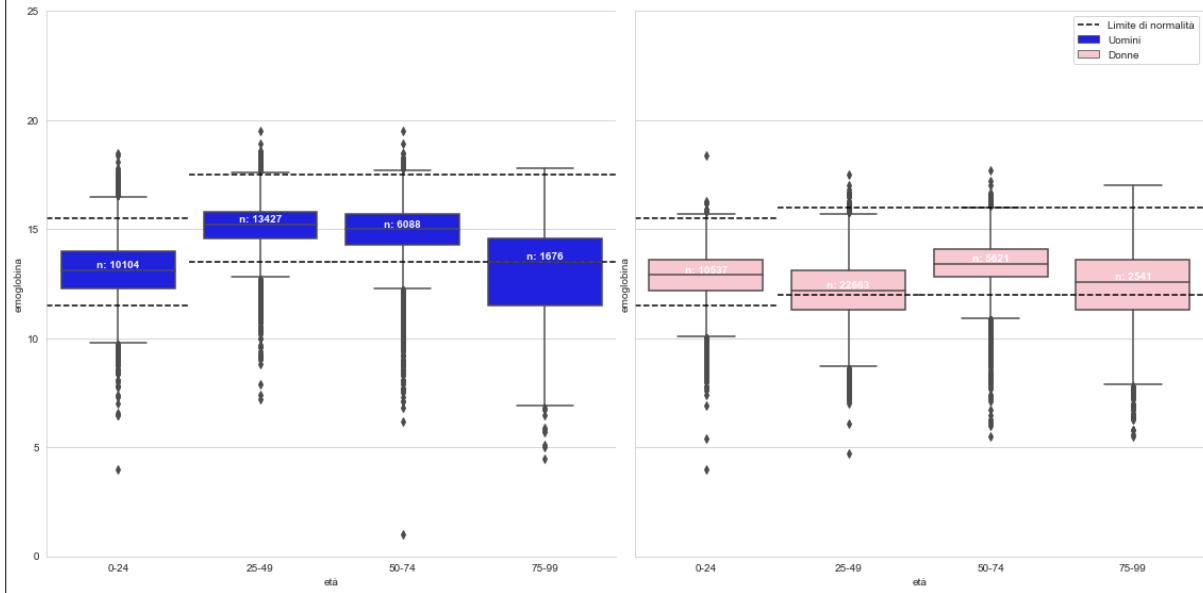
globuli rossi: popolazione completa (valori annui medi)
osservazioni totali: 2706420



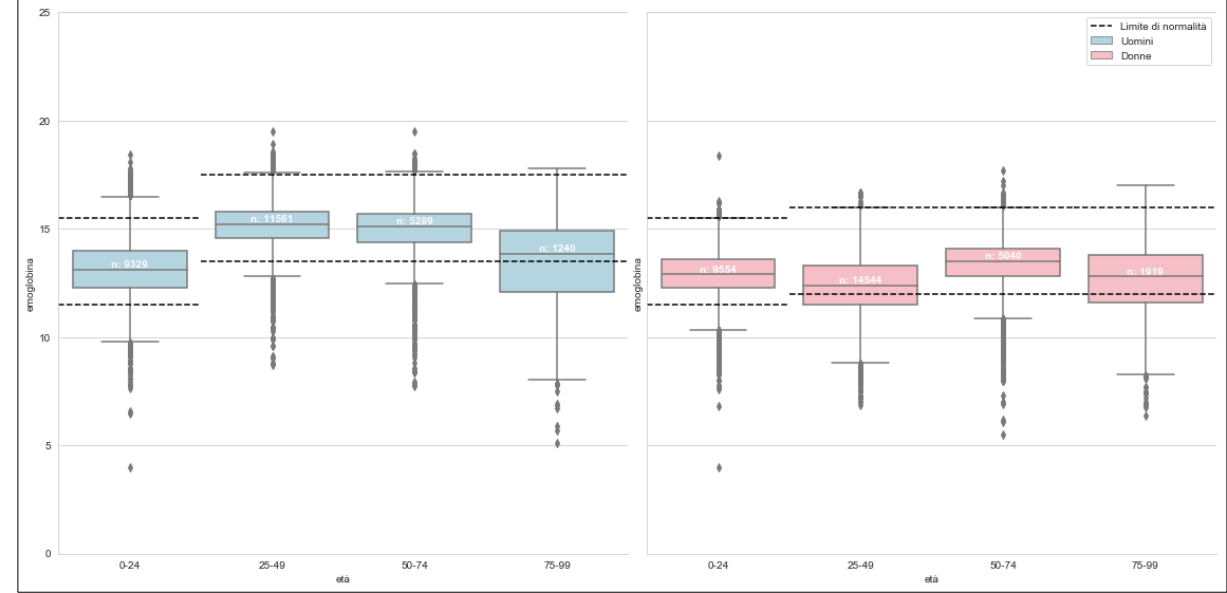
Fase 3: definizione sani

- Esclusione di tutte le persone che nell'arco temporale preso in esame abbiano avuto almeno una volta valori di glucosio, GPT o creatinina fuori dai parametri. Queste analisi rilevano infatti i più diffusi disturbi cronici che potrebbero inficiare i risultati delle analisi del sangue.
- Questa esclusione è una condizione molto forte che elimina dal campione una parte importante della popolazione non cronica ma che assicura che il campione selezionato sia composto certamente da persone “**sane**”, ovvero non affetti da patologie croniche.
- Il campione si è quindi ridotto a circa **sessantamila** osservazioni.
- **L'analisi di questo nuovo campione, anche valutando la media annua degli esiti analitici per singolo individuo, evidenzia un andamento in linea con i valori osservabili su tutto il campione preso in esame originariamente.**

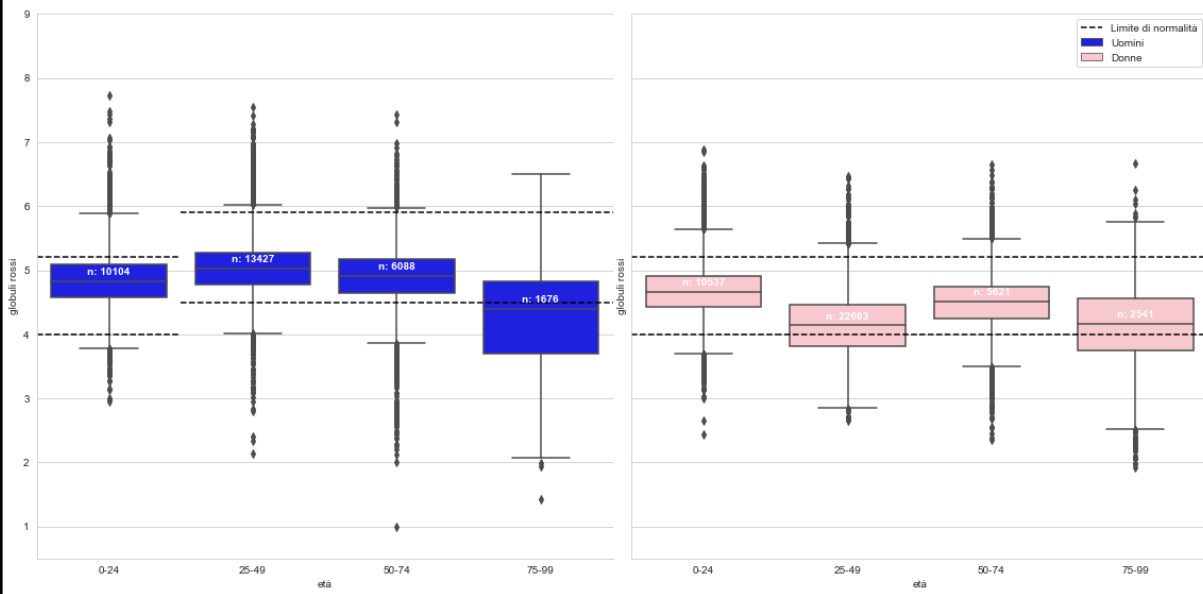
emoglobina: popolazione sana
osservazioni totali: 75022



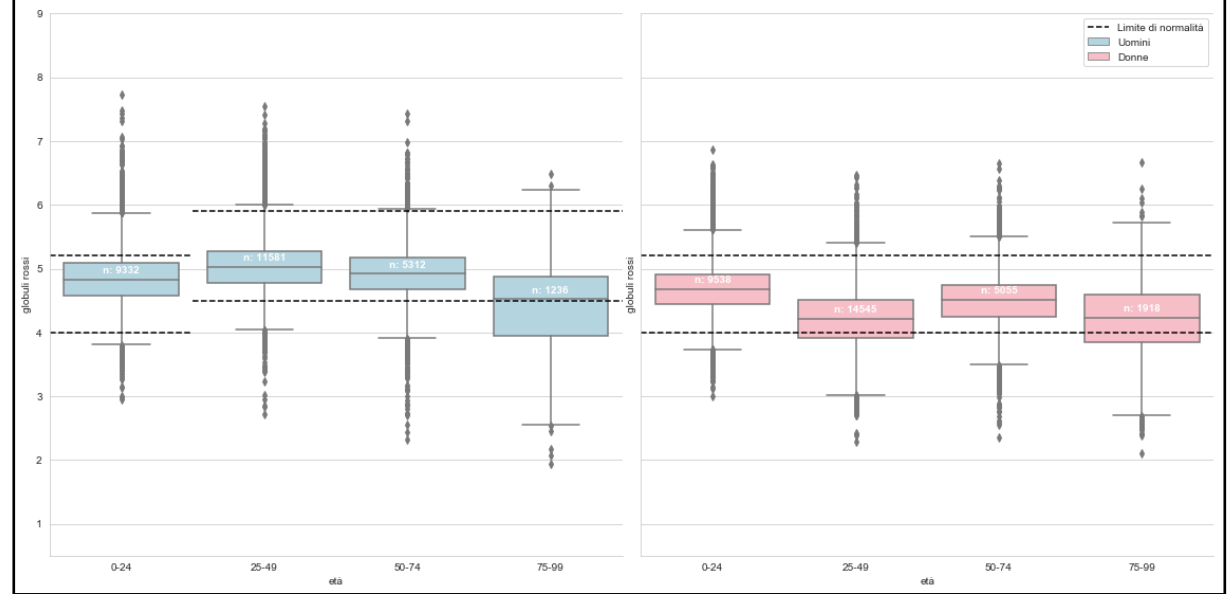
emoglobina: popolazione sana (valori annui medi).
osservazioni totali: 60433



globuli rossi: popolazione sana
osservazioni totali: 75022



globuli rossi: popolazione sana (valori annui medi).
osservazioni totali: 60474



Fase 4: la verifica della popolazione sana rispetto ad outcome clinici

La verifica del campione di popolazione considerata sana è stato effettuato utilizzando lo strumento RiskER messo a disposizione dalla Regione

Lo strumento RiskER rappresenta la popolazione residente stratificata in relazione al rischio di ospedalizzazione per cause prevenibili e decesso

Del cluster di popolazione sana precedentemente osservato è stato estratto quello con osservazioni fino al 31-12-2018, che alla data fosse maggiore di 14 anni, per poterlo confrontare con la classificazione RiskER del 2018

RiskER: le banche dati

- Anagrafe sanitaria per l'identificazione della popolazione in analisi, le caratteristiche demografiche e lo stato in vita
- Archivio delle schede di dimissione ospedaliera (SDO) per identificare i problemi di salute dei soggetti in anagrafe (codici diagnosi ICD9-CM) e gli accessi in ospedale
- Archivio degli accessi in Pronto soccorso (flusso PS) per identificare i problemi di salute dei soggetti in anagrafe (codici diagnosi ICD9-CM) e gli accessi in Pronto soccorso
- Archivio dei pazienti in assistenza domiciliare integrata (flusso ADI) per identificare i problemi di salute dei soggetti in anagrafe (codici diagnosi ICD9-CM)
- Archivio dell'assistenza specialistica ambulatoriale (flusso ASA) per identificare la presenza di malattia tumorale (T1-radioterapia/chemioterapia), malattie genitourinarie (99.25) e dialisi (T2-dialisi)
- Prescrizioni farmaceutiche dei soggetti in anagrafe (codice ATC): Archivio farmaceutica territoriale (AFT) e Archivio distribuzione diretta farmaci e distribuzione per conto (FED)

RiskER: il modello di previsione

Il modello di previsione del rischio è stato stimato sul totale della popolazione adulta (**≥14 anni**) **residente** in Emilia-Romagna al 31/12/2014 (quasi 4 milioni di persone), usando i dati sull'utilizzo dei servizi **sanitari (SDO, ADI, PS)**, i dati sulla **farmaceutica (AFT FED)** e i **dati demografici relativi ad età, sesso e geolocalizzazione geografica** nello stesso anno.

Per ogni singolo paziente è stato anche considerato **l'utilizzo dei servizi sanitari e i dati sulla farmaceutica nei cinque anni precedenti, cioè nel periodo 2009-2013.**

Come variabile dipendente è **stata utilizzata l'ospedalizzazione in regime ordinario e il decesso nell'anno 2015.**

Il modello è stato sviluppato utilizzando una regressione logistica su strati di popolazione.

18137 persone di età maggiore di 14 anni con esami di laboratorio fino al 31/12/2018 considerati sani dal punto di vista diagnostico

17753
appartenenti
alla classe
RiskER 1

384
appartenenti
alle classi
RiskER da 2 a
4

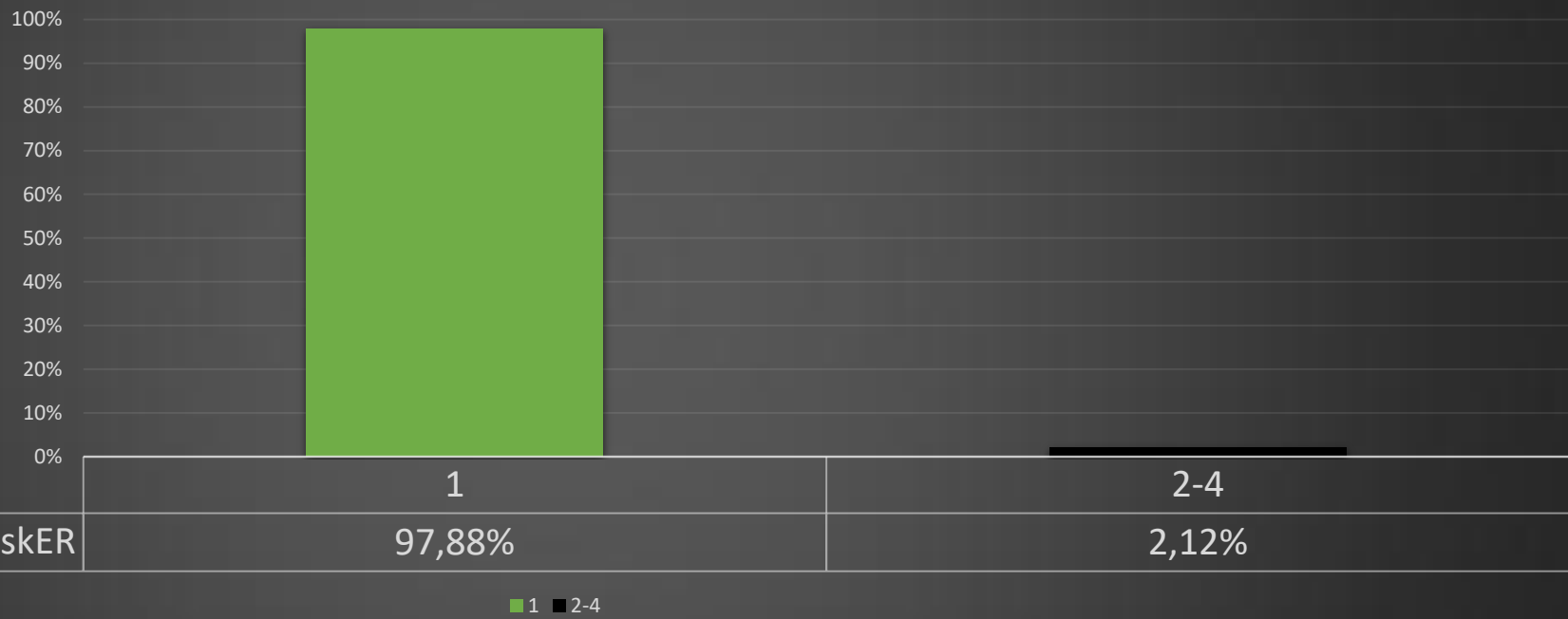
Distribuzione tra le classi a maggior rischio

284 classe 2

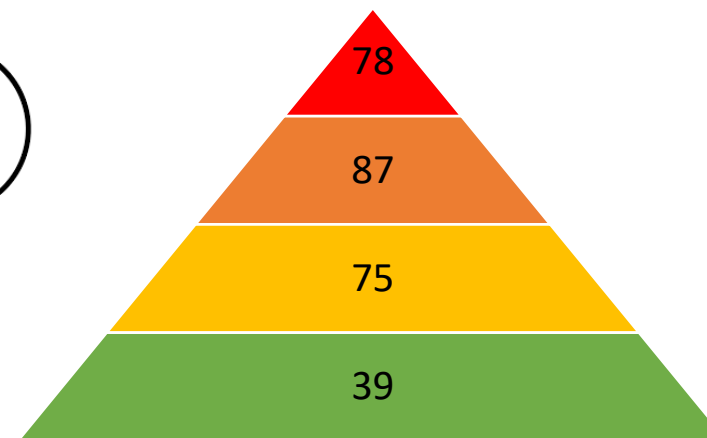
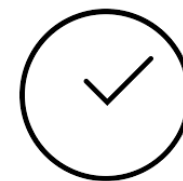
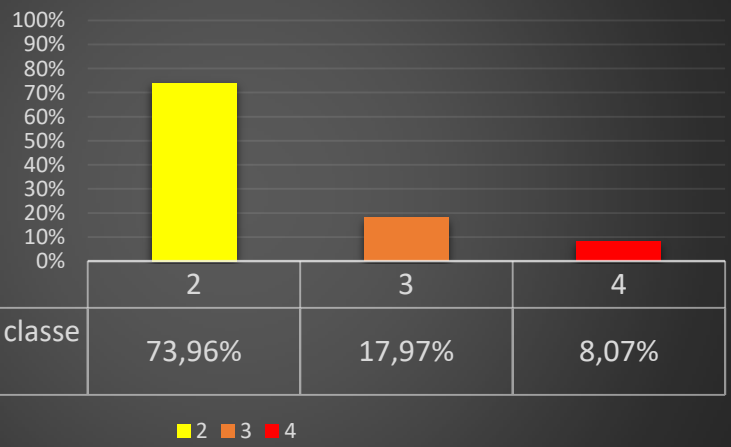
69 in classe 3

31 in classe 4

Percentuale per classe RiskER



Distribuzione delle classi da 2 a 4



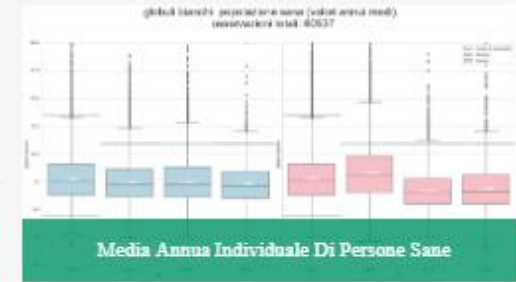
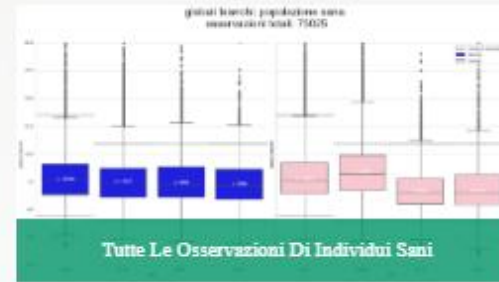
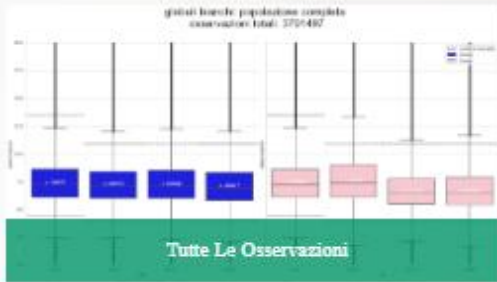
Età media del campione di popolazione sana
Per classificazione su fascia RiskER

Gli strumenti

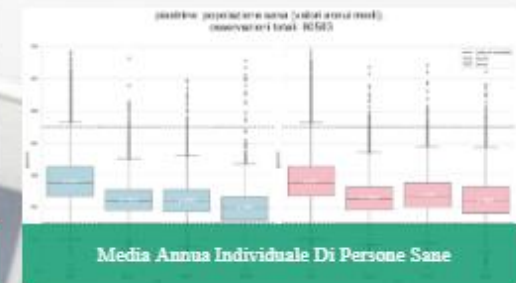
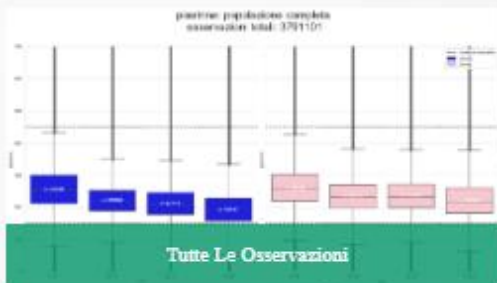
Per rispondere alle complesse e peculiari esigenze analitiche che si sono affrontate, l'AUSL di Modena ha predisposto uno specifico ambiente dedicato:

- A partire dal data warehouse aziendale è stato predisposto un datamart anonimizzato che consente il rispetto della privacy in accordo con le norme vigenti. Tale ambiente è isolato dall'ambiente di sviluppo aziendale e espone esclusivamente i dati utili per l'indagine.
- L'accesso ai dati è possibile esclusivamente all'interno del perimetro della rete aziendale
- È stata predisposto un ambiente (hardware e software) dedicato allo strumento di analisi statistica (Python 3.7), opportunamente dimensionato e integrato con l'infrastruttura di business intelligence aziendale
- **Questa modalità di lavoro ha permesso**
 - una buona libertà di analisi all'interno del perimetro stabilito,
 - ottima capacità di elaborazione, grazie ad una macchina remota dotata di specifiche elevate,
 - Possibilità di mantenere le informazioni in un ambiente protetto e controllato.

Globuli Bianchi



Piastrine



Definizione delle prestazioni sentinella

Emoglobina glicata (IFCC)

Creatinina

eGFR

Albumina urine

Proteine totali urine 24h

Glucosio

Emocromo/Emoglobina

Albumina

Proteine totali



Definizione dei criteri strutturati di validità per l'interpretazione del risultato

Unità di misura

Criteri di lettura del risultato numerico (come da referto)

Data inizio e fine validità

Selezione degli attributi anagrafici e amministrativi

Dati anagrafici paziente

Dati amministrativi prestazione

Dati anagrafici strutture (richiedenti ed eroganti)

Definizione dei range che contribuiscono alla composizione dei cluster di popolazione da analizzare

Valore ottimale

Valori indicativi

Soglie di patologia

End point ottimale in relazione alla terapia farmacologica e agli eventi critici

Valore anomalo

Valore grave

Fase 5: valutazione dell'impatto economico del comportamento anomalo fuori soglia

- Per la popolazione sana individuata identificare la prima data di insorgenza del fuori soglia rispetto alle prestazioni emoglobina, numero di globuli rossi, di globuli bianchi, di piastrine e volume medio globulare
- Verifica del consumo di prestazioni specialistiche e diagnostiche da flusso ASA dopo la data di insorgenza del fuori soglia, che non siano mai state consumate prima delle date di insorgenza del fuori soglia o in quantità inferiori al 25%
- Verifica del consumo di farmaci da flussi FED e AFT dopo la data di insorgenza del fuori soglia, che non siano mai state consumate prima delle date di insorgenza del fuori soglia o in quantità inferiori al 25%
- Quantificazione economica del consumo

Grazie per l'attenzione

t.trenti@ausl.mo.it

d.angeletti@ausl.mo.it

t.pirotti@ausl.mo.it