



# Il Progetto MATTONI/MATRICE dell'AGENAS

---

Rosa Gini

Agenzia regionale di sanità della Toscana, Firenze  
Dipartimento di Informatica Medica, Università Erasmus, Rotterdam



# Indice

Introduzione

Obiettivi e struttura del progetto MATRICE

Software

TheMatrix

Morpheus

TheOne

Gli studi di validazione

Gold Standard Validation (GSV)

Automatic Algorithm Validation (AAV)

Novel Algorithm Generation (NAG)

Manual Algorithm Validation (MAV)

Concordanza nella misura degli indicatori

TheMatrix oltre MATRICE

Discussione

# Indice

Introduzione

Obiettivi e struttura del progetto MATRICE

Software

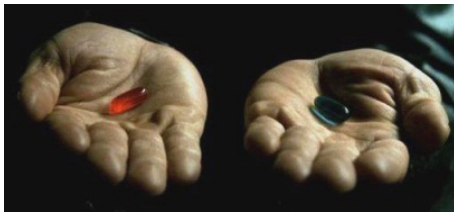
Gli studi di validazione

TheMatrix oltre MATRICE

Discussione

## Un monito

Un ricercatore di fronte alla vastità dei dati oggi disponibili fronteggia un dilemma.



## Un monito

Credere nei dati senza porsi  
domande...

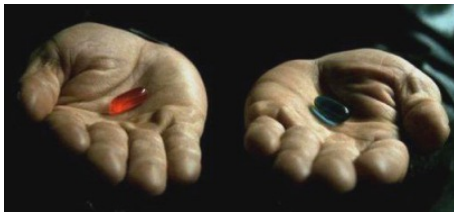


## Un monito




...o metterli alla prova per  
verificare la loro corrispondenza  
con la realtà

## Un monito



## Un monito



If you did not conduct your own validation study, be prepared to cite others who did. Validation studies increase cost and take time, but they may be the difference between cranking out analyses and sound epidemiologic research.

Hernán M. With great data comes great responsibility. *Epidemiology* 2011 May.



# Indice

Introduzione

Obbiettivi e struttura del progetto MATRICE

Software

Gli studi di validazione

TheMatrix oltre MATRICE

Discussione

## Obiettivi generali

... condividere basi informative comuni per arrivare ad un sistema che legga trasversalmente i sistemi sanitari regionali, rispetto a percorsi assistenziali articolati e continuativi. . .

dal documento di progetto

## Obiettivi generali

... utilizzare unicamente flussi informativi già esistenti, per poter osservare la gestione delle patologie complesse nelle diverse Regioni e Province Autonome, sulla base di dati la cui validità è condivisa con esse, attraverso operazioni di trasformazione del dato la cui validità verrà stabilita sulla base di rigorosi criteri scientifici. . .

dal documento di progetto

## Obiettivi generali

... al fine di contribuire ad evidenziare la misura dell'equità nell'erogazione delle cure territoriali ...

dal documento di progetto

## Obiettivi generali

... dotare i vari livelli istituzionali di una base conoscitiva condivisa che supporti la *governance* delle cure territoriali ...

dal documento di progetto

## Obiettivi specifici

Il progetto prevede il disegno, la sperimentazione, la validazione e l'applicazione ad alcuni casi di studio di un software open source di aggregazione e di integrazione dei dati dei flussi sanitari nazionali, per produrre la lettura dei percorsi assistenziali di pazienti con patologie complesse

dal documento di progetto

# Struttura

- P1 - **Patologie-Modelli informativi** Selezione delle patologie, enumerazione di algoritmi esistenti, validazione, generazione algoritmi innovativi, scelta, responsabile: ARS
- P2 - **Sistemi di indicatori** Selezione, calcolo e validazione di indicatori di adesione a percorsi assistenziali, responsabile: Agenas
- P3 - **Data Integration** Contesto, sviluppo e sperimentazione del processo di erogazione dati, responsabile: Ministero (NSIS)
- P4 - **Sperimentazione** Modelli di analisi dei flussi di dati secondari, sperimentazione della loro utilità pratica nell'organizzazione del servizio sanitario a livello locale e centrale, responsabile: Agenas
- P5 - **Management e comunicazione** responsabile: Agenas

## Partner

- Agenzia nazionale per i servizi sanitari regionali (Agenas)
- 5 Regioni (6 ASL + 1 Regione)
- Ministero della Salute
- ARS Toscana
- ISTI-CNR
- SIMG
- Dipartimento di Informatica Medica dell'Università Erasmus di Rotterdam (EMC)



# Tempi

febbraio 2011-febbraio 2013 (proroga 1 anno?)

# Primo prodotto Pacchetto 1

## Diabete

- Non complicato (comp/scomp)
- Danno d'organo asintomatico (c/s)
- Complicato (c/s)

## Cardiopatia ischemica

- Coronaropatia
- IMA
- Disfunzione ventricolare sinistra
- Insufficienza cardiaca

## Ipertensione

- Non complicata
- Complicata, no insufficienza cardiaca
- Insufficienza cardiaca

## Insufficienza cardiaca

- Stadio C

# Indice

Introduzione

Obiettivi e struttura del progetto MATRICE

Software

TheMatrix

Morpheus

TheOne

Gli studi di validazione

TheMatrix oltre MATRICE

Discussione

# Indice

Introduzione

Obbiettivi e struttura del progetto MATRICE

Software

TheMatrix

Morpheus

TheOne

Gli studi di validazione

TheMatrix oltre MATRICE

Discussione

## Data management di dati amministrativi

- Si installa presso un DBMS contenente flussi amministrativi (ASL, Regione)

## Data management di dati amministrativi

- Si installa presso un DBMS contenente flussi amministrativi (ASL, Regione)
- Funzionalità principale: esegue operazioni di data management complesse

## Data management di dati amministrativi

- Si installa presso un DBMS contenente flussi amministrativi (ASL, Regione)
- Funzionalità principale: esegue operazioni di data management complesse
- **Sviluppato dall'ISTI/CNR per MATRICE**

## Data management di dati amministrativi

- Si installa presso un DBMS contenente flussi amministrativi (ASL, Regione)
- Funzionalità principale: esegue operazioni di data management complesse
- Sviluppato dall'ISTI/CNR per MATRICE
- Estende Jerboa, sviluppato dall'EMC per EU-ADR



## Data management di dati amministrativi

- Si installa presso un DBMS contenente flussi amministrativi (ASL, Regione)
- Funzionalità principale: esegue operazioni di data management complesse
- Sviluppato dall'ISTI/CNR per MATRICE
- Estende Jerboa, sviluppato dall'EMC per EU-ADR
- Java

## Data management di dati amministrativi

- Si installa presso un DBMS contenente flussi amministrativi (ASL, Regione)
- Funzionalità principale: esegue operazioni di data management complesse
- Sviluppato dall'ISTI/CNR per MATRICE
- Estende Jerboa, sviluppato dall'EMC per EU-ADR
- Java
- Rilasciato con licenza GPL durante il progetto

## Funzionalità

- Interfacciamento a più sorgenti dati (csv, DBMS)

## Funzionalità

- Interfacciamento a più sorgenti dati (csv, DBMS)
- Flessibilità di configurazione

## Funzionalità

- Interfacciamento a più sorgenti dati (csv, DBMS)
- Flessibilità di configurazione
- Esecuzione automatizzata, in-place, di algoritmi di analisi

## Funzionalità

- Interfacciamento a più sorgenti dati (csv, DBMS)
- Flessibilità di configurazione
- Esecuzione automatizzata, in-place, di algoritmi di analisi
- Estrazione di indicatori e dataset

## Funzionalità

- Interfacciamento a più sorgenti dati (csv, DBMS)
- Flessibilità di configurazione
- Esecuzione automatizzata, in-place, di algoritmi di analisi
- Estrazione di indicatori e dataset
- Meccanismi sicuri di distribuzione degli script

## Funzionalità

- Interfacciamento a più sorgenti dati (csv, DBMS)
- Flessibilità di configurazione
- Esecuzione automatizzata, in-place, di algoritmi di analisi
- Estrazione di indicatori e dataset
- Meccanismi sicuri di distribuzione degli script
- Criptazione tramite chiave da ricevere da webservice



## Funzionalità

- Interfacciamento a più sorgenti dati (csv, DBMS)
- Flessibilità di configurazione
- Esecuzione automatizzata, in-place, di algoritmi di analisi
- Estrazione di indicatori e dataset
- Meccanismi sicuri di distribuzione degli script
- Criptazione tramite chiave da ricevere da webservice
- Sicurezza, privacy dei trasferimenti di dati

# Installazione

- Applicazione Java: portabile

# Installazione

- Applicazione Java: portabile
- Distribuita via secure ftp

# Installazione

- Applicazione Java: portabile
- Distribuita via secure ftp
- Adattare la configurazione al sistema via file XML

# Installazione

- Applicazione Java: portabile
- Distribuita via secure ftp
- Adattare la configurazione al sistema via file XML
  - Path dei dati, connessione a DBMS, utenti del sistema

# Installazione

- Applicazione Java: portabile
- Distribuita via secure ftp
- Adattare la configurazione al sistema via file XML
  - Path dei dati, connessione a DBMS, utenti del sistema
  - Schema dei dati disponibili localmente

## Attualmente

- Sviluppato sul datawarehouse dell'ARS

## Attualmente

- Sviluppato sul datawarehouse dell'ARS
- Marzo: rilasciato un prototipo installato in 7 aziende



# Indice

Introduzione

Obbiettivi e struttura del progetto MATRICE

**Software**

TheMatrix

**Morpheus**

TheOne

Gli studi di validazione

TheMatrix oltre MATRICE

Discussione

## Plug-in per software clinico MMG

- Si installa sul software Millewin di un MMG

## Plug-in per software clinico MMG

- Si installa sul software Millewin di un MMG
- Funzionalità principale: esegue operazioni di data management complesse

## Plug-in per software clinico MMG

- Si installa sul software Millewin di un MMG
- Funzionalità principale: esegue operazioni di data management complesse
- Sviluppato da risorse della SIMG su disegno di MATRICE

## Plug-in per software clinico MMG

- Si installa sul software Millewin di un MMG
- Funzionalità principale: esegue operazioni di data management complesse
- Sviluppato da risorse della SIMG su disegno di MATRICE
- Software commerciale

# Indice

Introduzione

Obbiettivi e struttura del progetto MATRICE

Software

TheMatrix

Morpheus

TheOne

Gli studi di validazione

TheMatrix oltre MATRICE

Discussione

## Reporting

- Si installa nello stesso sistema che ospita TheMatrix

# Reporting

- Si installa nello stesso sistema che ospita TheMatrix
- Funzionalità principale: reporting



# Reporting

- Si installa nello stesso sistema che ospita TheMatrix
- Funzionalità principale: reporting
- Sviluppato da Agenas

# Reporting

- Si installa nello stesso sistema che ospita TheMatrix
- Funzionalità principale: reporting
- Sviluppato da Agenas
- R

## Reporting

- Si installa nello stesso sistema che ospita TheMatrix
- Funzionalità principale: reporting
- Sviluppato da Agenas
- R
- Verrà rilasciato con licenza GPL durante il progetto

## Input per TheOne

Nel progetto TheMatrix genera un tracciato record che è l'input di TheOne

# Tracciato record individuale dell'input

Dati strutturali DATA, DATA\_AGG

# Tracciato record individuale dell'input

Dati strutturali DATA, DATA\_AGG

Dati generali sul soggetto ID\_SOGGETTO, DATA\_NASCITA, SESSO,  
CITTADINANZA, COMUNE, DENSITA\_COMUNE,  
INDIRIZZO, SEZIONE, INDICE\_DEPRIVAZ,  
DATA\_INGRESSO, DATA\_USCITA, DATA\_MORTE,  
IN\_CARICO, IN\_CARICO12M

# Tracciato record individuale dell'input

Dati strutturali DATA, DATA\_AGG

Dati generali sul soggetto ID\_SOGGETTO, DATA\_NASCITA, SESSO, CITTADINANZA, COMUNE, DENSITA\_COMUNE, INDIRIZZO, SEZIONE, INDICE\_DEPRIVAZ, DATA\_INGRESSO, DATA\_USCITA, DATA\_MORTE, IN\_CARICO, IN\_CARICO12M

Dati MMG MMG, DATA\_MMG, MMG\_12M

# Tracciato record individuale dell'input

Dati strutturali DATA, DATA\_AGG

Dati generali sul soggetto ID\_SOGGETTO, DATA\_NASCITA, SESSO, CITTADINANZA, COMUNE, DENSITA\_COMUNE, INDIRIZZO, SEZIONE, INDICE\_DEPRIVAZ, DATA\_INGRESSO, DATA\_USCITA, DATA\_MORTE, IN\_CARICO, IN\_CARICO12M

Dati MMG MMG, DATA\_MMG, MMG\_12M

Esenzioni e consumi NUM\_ESENZIONI, COSTO\_LORDO\_FARMACI, COSTO\_LORDO\_SPECIAL\_AMB, COSTO\_LORDO\_OSP, COSTO\_LORDO\_ALTRO, NUM\_PRESC\_FARM, NUM\_RIC



# Tracciato record individuale dell'input

Dati strutturali DATA, DATA\_AGG

Dati generali sul soggetto ID\_SOGGETTO, DATA\_NASCITA, SESSO, CITTADINANZA, COMUNE, DENSITA\_COMUNE, INDIRIZZO, SEZIONE, INDICE\_DEPRIVAZ, DATA\_INGRESSO, DATA\_USCITA, DATA\_MORTE, IN\_CARICO, IN\_CARICO12M

Dati MMG MMG, DATA\_MMG, MMG\_12M

Esenzioni e consumi NUM\_ESENZIONI, COSTO\_LORDO\_FARMACI, COSTO\_LORDO\_SPECIAL\_AMB, COSTO\_LORDO\_OSP, COSTO\_LORDO\_ALTRO, NUM\_PRESC\_FARM, NUM\_RIC

Diabete DATA\_DIAGN\_DIABETE, STADIO\_DIABETE, STADIO\_DIABETE\_12M, NUM\_INDICATORE\_DIABETE\_1, ..., NUM\_INDICATORE\_DIABETE\_n, DENOM\_INDICATORE\_DIABETE\_1, ..., DENOM\_INDICATORE\_DIABETE\_n, ESENZIONE\_DIABETE

# Tracciato record individuale dell'input

Dati strutturali DATA, DATA\_AGG

Dati generali sul soggetto ID\_SOGGETTO, DATA\_NASCITA, SESSO, CITTADINANZA, COMUNE, DENSITA\_COMUNE, INDIRIZZO, SEZIONE, INDICE\_DEPRIVAZ, DATA\_INGRESSO, DATA\_USCITA, DATA\_MORTE, IN\_CARICO, IN\_CARICO12M

Dati MMG MMG, DATA\_MMG, MMG\_12M

Esenzioni e consumi NUM\_ESENZIONI, COSTO\_LORDO\_FARMACI, COSTO\_LORDO\_SPECIAL\_AMB, COSTO\_LORDO\_OSP, COSTO\_LORDO\_ALTRO, NUM\_PRESC\_FARM, NUM\_RIC

Diabete DATA\_DIAGN\_DIABETE, STADIO\_DIABETE, STADIO\_DIABETE\_12M, NUM\_INDICATORE\_DIABETE\_1, ..., NUM\_INDICATORE\_DIABETE\_n, DENOM\_INDICATORE\_DIABETE\_1, ..., DENOM\_INDICATORE\_DIABETE\_n, ESENZIONE\_DIABETE

Ipertensione ...

# Indice

Introduzione

Obiettivi e struttura del progetto MATRICE

Software

**Gli studi di validazione**

Gold Standard Validation (GSV)

Automatic Algorithm Validation (AAV)

Novel Algorithm Generation (NAG)

Manual Algorithm Validation (MAV)

Concordanza nella misura degli indicatori

TheMatrix oltre MATRICE

# Indice

Introduzione

Obiettivi e struttura del progetto MATRICE

Software

**Gli studi di validazione**

Gold Standard Validation (GSV)

Automatic Algorithm Validation (AAV)

Novel Algorithm Generation (NAG)

Manual Algorithm Validation (MAV)

Concordanza nella misura degli indicatori

TheMatrix oltre MATRICE

## Schema di GSV

Da validare algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG candidati a essere il Gold Standard di AAV

## Schema di GSV

Da validare algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG  
candidati a essere il Gold Standard di AAV

Gold standard validazione manuale del MMG

## Schema di GSV

**Da validare** algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG candidati a essere il Gold Standard di AAV

**Gold standard** validazione manuale del MMG

**Data collection** semiautomatica: Morpheus estrae i dati dai database clinici dei MMG e propone un'interfaccia di inserimento dati che i MMG compilano a mano

## Schema di GSV

**Da validare** algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG candidati a essere il Gold Standard di AAV

**Gold standard** validazione manuale del MMG

**Data collection** semiautomatica: Morpheus estrae i dati dai database clinici dei MMG e propone un'interfaccia di inserimento dati che i MMG compilano a mano

**Attori** SIMG (principal investigator, sviluppo di Morpheus e reclutamento MMG), ARS (responsabile e principal investigator a ugual contributo)



# Estrazione automatica, il MMG non vi ha accesso

nome variabile	significato	formato
cod_fis_paz	codice fiscale del paziente	
nome_paz	nome del paziente	
cognome_paz	cognome del paziente	
sesto	genere del paziente	
quinquennio_nascita	quinquennio di nascita del paziente	
iperten0	se il paziente risulta iperteso	binaria
giorni_iperten0	data_indice - data dalla quale il paziente risulta iperteso	intero
iperten1	se il paziente risulta in stadio 1 dell'ipertensione	binaria
giorni_iperten1	data_indice - data dalla quale il paziente risulta in stadio 1 dell'ipertensione	intero
iperten2	se il paziente risulta in stadio 2 dell'ipertensione	binaria
giorni_iperten2	data_indice - data dalla quale il paziente risulta in stadio 2 dell'ipertensione	intero
iperten3	se il paziente risulta in stadio 3 dell'ipertensione	binaria
giorni_iperten3	data_indice - data dalla quale il paziente risulta in stadio 3 dell'ipertensione	intero
diab0	se il paziente risulta diabetico	binaria
giorni_diab0	data_indice - data dalla quale il paziente risulta diabetico	intero
diab1a	se il paziente risulta in stadio 1a del diabete	binaria
giorni_diab1a	data_indice - data dalla quale il paziente risulta in stadio 1a del diabete	intero

...

# Interfaccia per la compilazione manuale del MMG (circa 200 paz)

<b>Nome</b>	<b>Cognome</b>	<b>Quinquennio di nascita</b>	<b>Ipertensione</b>	<b>Diabete</b>	<b>Cardiopatìa Ischemica</b>	<b>Insufficienza cardiaca</b>	<b>Note</b>
Elena	Rossi	1967-1972					
Francesco	Bianchi	1942-1947					
Loredana	Verdi	1927-1932					
Paolo	Romeni	1942-1947					
Emanuela	Aqui	1947-1952					
Lorenzo	Franchi	1972-1977					
...	...	...					

## Criticità di GSV

- I MMG che partecipano a GSV sono particolarmente motivati

## Criticità di GSV

- I MMG che partecipano a GSV sono particolarmente motivati
- Potrebbero non essere rappresentativi dei medici del database della SIMG, in particolare di quelli reclutati per AAV

## Criticità di GSV

- I MMG che partecipano a GSV sono particolarmente motivati
- Potrebbero non essere rappresentativi dei medici del database della SIMG, in particolare di quelli reclutati per AAV
- Eseguiamo uno studio di sensibilità sul database della SIMG per stabilire le differenze di profilo di registrazione tra i medici campione e gli altri

# Indice

Introduzione

Obiettivi e struttura del progetto MATRICE

Software

**Gli studi di validazione**

Gold Standard Validation (GSV)

Automatic Algorithm Validation (AAV)

Novel Algorithm Generation (NAG)

Manual Algorithm Validation (MAV)

Concordanza nella misura degli indicatori

TheMatrix oltre MATRICE

## AAV: lo studio cuore di MATRICE

- L'uso dei dati amministrativi applicato in modo sistematico al governo delle patologie croniche deve essere fondato su uno studio di validazione

## AAV: lo studio cuore di MATRICE

- L'uso dei dati amministrativi applicato in modo sistematico al governo delle patologie croniche deve essere fondato su uno studio di validazione
- Per questo MATRICE prevede al proprio cuore la validazione degli algoritmi di riconoscimento di caso



## AAV: lo studio cuore di MATRICE

- L'uso dei dati amministrativi applicato in modo sistematico al governo delle patologie croniche deve essere fondato su uno studio di validazione
- Per questo MATRICE prevede al proprio cuore la validazione degli algoritmi di riconoscimento di caso
- Lo studio AAV permetterà di descrivere le caratteristiche degli algoritmi di riconoscimento di caso esistenti e di scegliere tra essi gli algoritmi migliori per gli scopi del progetto

## Schema di AAV

Da validare un ampio spettro di algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

## Schema di AAV

Da validare un ampio spettro di algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

Gold standard algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG

## Schema di AAV

**Da validare** un ampio spettro di algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

**Gold standard** algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG

**Data collection** completamente automatica: TheMatrix estrae i dati dai database amministrativi aziendali e Morpheus estrae i dati dai database clinici dei MMG

## Schema di AAV

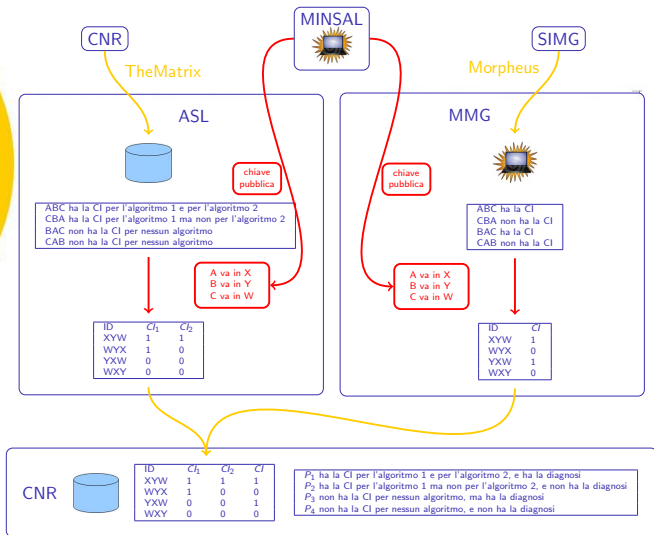
**Da validare** un ampio spettro di algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

**Gold standard** algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG

**Data collection** completamente automatica: TheMatrix estrae i dati dai database amministrativi aziendali e Morpheus estrae i dati dai database clinici dei MMG

**Attori** ARS (responsabile e principal investigator), CNR (sviluppo di TheMatrix), SIMG (supporto scientifico, sviluppo di Morpheus e reclutamento MMG), MINSAN (privacy), AGENAS (etica), aziende ASL (installazione e lancio di TheMatrix)

# Raccolta dati



## Assunzioni in AAV

- ① Un medico di base è a conoscenza delle patologie croniche delle quali sono affetti i suoi pazienti

## Assunzioni in AAV

- ① Un medico di base è a conoscenza delle patologie croniche delle quali sono affetti i suoi pazienti
- ② I medici che partecipano al database della SIMG sono dei *buoni codificatori*, quindi registrano in modo omogeneo e uniforme le informazioni riguardanti le patologie croniche dei pazienti



## Assunzioni in AAV

- ① Un medico di base è a conoscenza delle patologie croniche delle quali sono affetti i suoi pazienti
- ② I medici che partecipano al database della SIMG sono dei *buoni codificatori*, quindi registrano in modo omogeneo e uniforme le informazioni riguardanti le patologie croniche dei pazienti
- ③ Gli algoritmi di riconoscimento di caso automatico dai database clinici della medicina generale validati da GSV sono solidi

## Assunzioni in AAV

- 1 Un medico di base è a conoscenza delle patologie croniche delle quali sono affetti i suoi pazienti
- 2 I medici che partecipano al database della SIMG sono dei *buoni codificatori*, quindi registrano in modo omogeneo e uniforme le informazioni riguardanti le patologie croniche dei pazienti
- 3 Gli algoritmi di riconoscimento di caso automatico dai database clinici della medicina generale validati da GSV sono solidi
- 4 Gli assistiti dei medici che partecipano al database della SIMG sono simili agli altri in quanto a percorsi diagnostici (verificabile)

# Indice

Introduzione

Obiettivi e struttura del progetto MATRICE

Software

**Gli studi di validazione**

Gold Standard Validation (GSV)

Automatic Algorithm Validation (AAV)

Novel Algorithm Generation (NAG)

Manual Algorithm Validation (MAV)

Concordanza nella misura degli indicatori

TheMatrix oltre MATRICE

## Descrizione di NAG

- Scopo dello studio è generare algoritmi basati su dati amministrativi e predittivi della condizione di patologia del soggetto

## Descrizione di NAG

- Scopo dello studio è generare algoritmi basati su dati amministrativi e predittivi della condizione di patologia del soggetto
- Lo studio verrà condotto tramite l'uso di strumenti di apprendimento automatico

## Schema di NAG

Da generare nuovi algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

## Schema di NAG

Da generare nuovi algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

Gold standard algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG

## Schema di NAG

**Da generare** nuovi algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

**Gold standard** algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG

**Data collection** contemporanea all'estrazione di dati di AAV



## Schema di NAG

- Da generare nuovi algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi
- Gold standard algoritmi di riconoscimento di caso basati su dati clinici della medicina generale ed estratti da database di medici partecipanti al database della SIMG
- Data collection contemporanea all'estrazione di dati di AAV
  - Attori CNR (responsabile e principal investigator), ARS e SIMG (supporto scientifico), MINSAN (privacy), AGENAS (etica)

## Razionale di NAG

- Gli strumenti di apprendimento automatico sono spesso in grado di generare algoritmi con altissime specificità e sensibilità

## Razionale di NAG

- Gli strumenti di apprendimento automatico sono spesso in grado di generare algoritmi con altissime specificità e sensibilità
- In caso di successo, gli algoritmi generati possono essere studiati da un punto di vista causale e considerati come candidati alternativi agli algoritmi migliori individuati da AAV

# Indice

Introduzione

Obiettivi e struttura del progetto MATRICE

Software

**Gli studi di validazione**

Gold Standard Validation (GSV)

Automatic Algorithm Validation (AAV)

Novel Algorithm Generation (NAG)

**Manual Algorithm Validation (MAV)**

Concordanza nella misura degli indicatori

TheMatrix oltre MATRICE

## MAV: un'opportunità?

- Nel caso in cui ci sia un'intersezione tra i MMG partecipanti a AAV e i MMG partecipanti a GSV è possibile utilizzare la validazione manuale anche per validare gli algoritmi estratti dai database amministrativi

## MAV: un'opportunità?

- Nel caso in cui ci sia un'intersezione tra i MMG partecipanti a AAV e i MMG partecipanti a GSV è possibile utilizzare la validazione manuale anche per validare gli algoritmi estratti dai database amministrativi
- Probabilmente i numeri saranno troppo bassi per ottenere stime statisticamente significative, ma questo può essere un interessante studio di sensibilità di AAV

## Schema di MAV

Da validare un ampio spettro di algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

## Schema di MAV

Da validare un ampio spettro di algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

Gold standard validazione manuale del MMG



## Schema di MAV

Da validare un ampio spettro di algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

Gold standard validazione manuale del MMG

Data collection contestuale al data collection di GSV (con chiavi di criptazione) e AAV

## Schema di MAV

**Da validare** un ampio spettro di algoritmi di riconoscimento di caso (patologia + stadiazione) basati su dati amministrativi

**Gold standard** validazione manuale del MMG

**Data collection** contestuale al data collection di GSV (con chiavi di criptazione) e AAV

**Attori** ARS (responsabile e principal investigator), SIMG (supporto scientifico), CNR (sviluppo sistema di criptazione)

# Indice

Introduzione

Obiettivi e struttura del progetto MATRICE

Software

**Gli studi di validazione**

Gold Standard Validation (GSV)

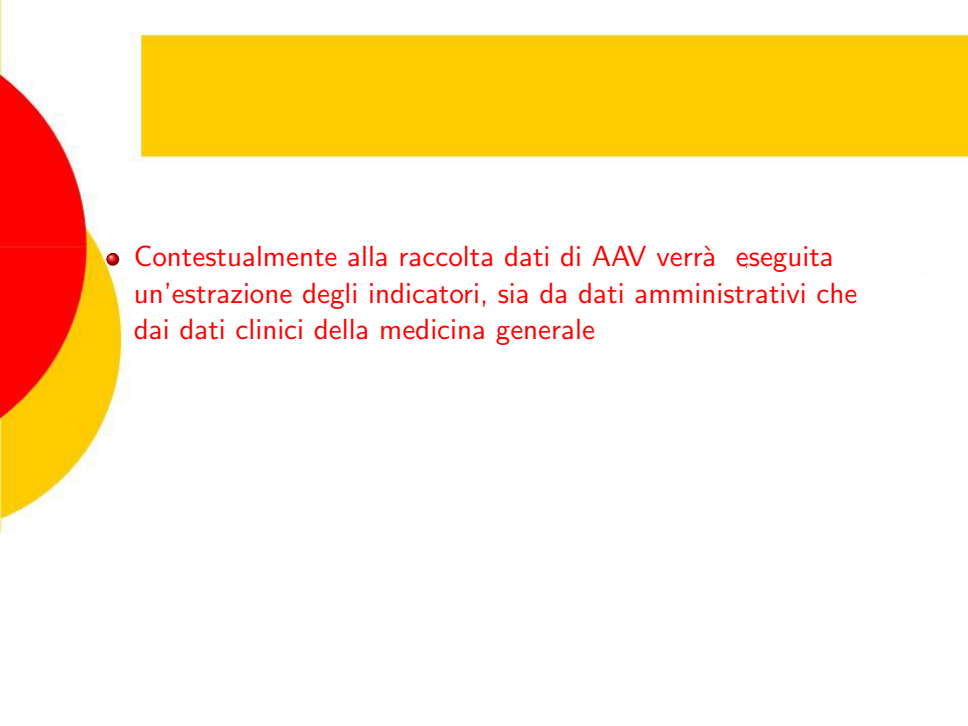
Automatic Algorithm Validation (AAV)

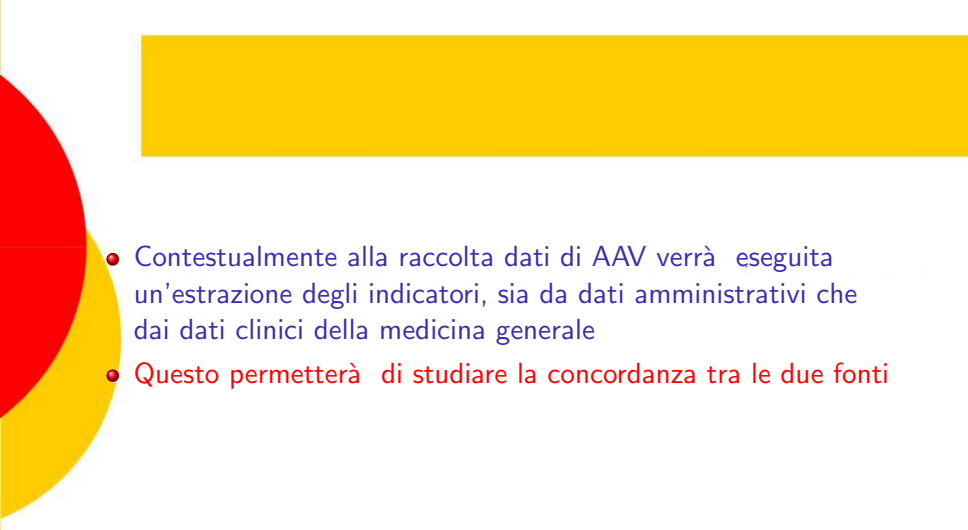
Novel Algorithm Generation (NAG)



Manual Algorithm Validation (MAV)

**Concordanza nella misura degli indicatori**

TheMatrix oltre MATRICE

- 
- Contestualmente alla raccolta dati di AAV verrà eseguita un'estrazione degli indicatori, sia da dati amministrativi che dai dati clinici della medicina generale

- 
- Contestualmente alla raccolta dati di AAV verrà eseguita un'estrazione degli indicatori, sia da dati amministrativi che dai dati clinici della medicina generale
  - Questo permetterà di studiare la concordanza tra le due fonti

- 
- 
- Contestualmente alla raccolta dati di AAV verrà eseguita un'estrazione degli indicatori, sia da dati amministrativi che dai dati clinici della medicina generale
  - Questo permetterà di studiare la concordanza tra le due fonti
  - Non si tratta quindi di validazione perché nessuna delle due fonti può essere considerata gold standard

# Indice

Introduzione

Obbiettivi e struttura del progetto MATRICE

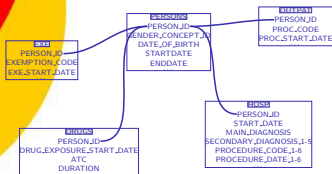
Software

Gli studi di validazione

**TheMatrix oltre MATRICE**

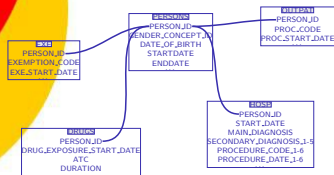
Discussione

# Raccolta dati di studi osservazionali da database amministrativi



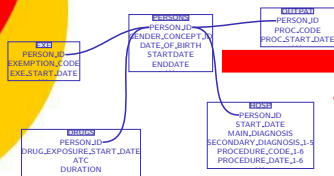


# Raccolta dati di studi osservazionali da database amministrativi



ANONYMOUS\_ID  
GENDER\_CONCEPT\_ID  
BIRTHYEAR  
EXPOSURE  
COVARIATE<sub>1</sub>  
...  
COVARIATE<sub>N</sub>  
RESPONSE

# Raccolta dati di studi osservazionali da database amministrativi



ANONYMOUS\_ID  
GENDER\_CONCEPT\_ID  
BIRTHYEAR  
EXPOSURE  
COVARIATE<sub>1</sub>  
...  
COVARIATE<sub>N</sub>  
RESPONSE

## Interfaccia grafica

- TheMatrix esegue degli script che generano dataset

## Interfaccia grafica

- TheMatrix esegue degli script che generano dataset
- Il linguaggio di scripting è estremamente flessibile

## Interfaccia grafica

- TheMatrix esegue degli script che generano dataset
- Il linguaggio di scripting è estremamente flessibile
- **Sviluppare uno script richiede competenze di programmazione**

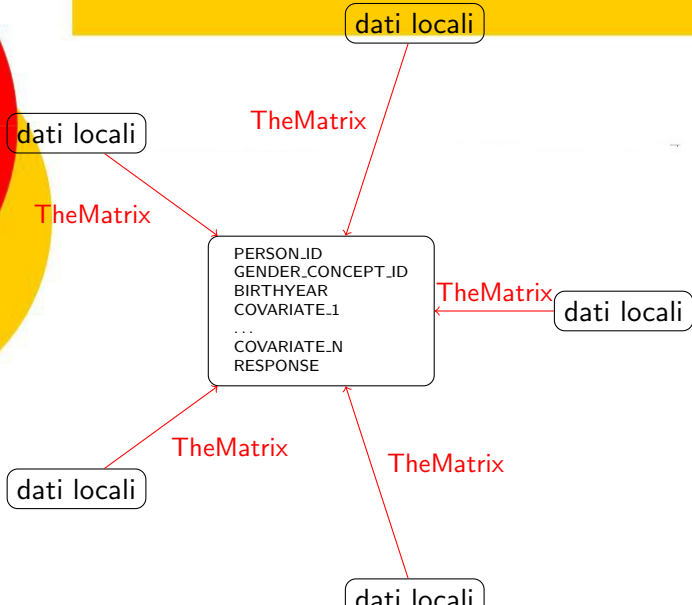
# Interfaccia grafica

- TheMatrix esegue degli script che generano dataset
- Il linguaggio di scripting è estremamente flessibile
- Sviluppare uno script richiede competenze di programmazione
- È in corso di sviluppo un'interfaccia grafica che supporta il ricercatore nella generazione degli script

## Interfaccia grafica

- TheMatrix esegue degli script che generano dataset
- Il linguaggio di scripting è estremamente flessibile
- Sviluppare uno script richiede competenze di programmazione
- È in corso di sviluppo un'interfaccia grafica che supporta il ricercatore nella generazione degli script
- Grazie all'interfaccia sviluppare uno script sarà maggiormente intuitivo

# Studi multicentrici





# Workshop a MIE2012: oltre i confini

## Frameworks for Data Extraction and Management from Electronic Healthcare Databases for Multi-Center Epidemiologic Studies: a Comparison among EU-ADR, MATRICE, and OMOP Strategies

Rosa GEN<sup>1\*</sup>, Massimo COPPOLA<sup>2</sup>, Patrick B RYAN<sup>3</sup>, Giacomo REGHETTI,  
Isauro PERE<sup>4</sup>, Roberto BERINI<sup>5</sup>, Paul AVULAGI<sup>6</sup>, Francisco M COLOMA<sup>7</sup>, Gianluca  
TRIFIRO<sup>8</sup>, Cayo DIALLO<sup>9</sup>, Johan VAN DER LEP, Miriam CIM

STURKENBOOM<sup>10</sup> and Marijn J SCHEMML<sup>11</sup>

<sup>1</sup> Agenzia regionale di sanità della Toscana, Florence, Italy

<sup>2</sup> Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands

<sup>3</sup> Istituto di Scienze e Tecnologie dell'Informazione "A. Fiaschi", National Research Council, Pisa, Italy

<sup>4</sup> Observational Medical Outcomes Partnership, USA

<sup>5</sup> HEGP Hospital, University Paris Descartes, Paris, France

<sup>6</sup> Department of Clinical and Experimental Medicine and Pharmacology, Section of Pharmacology, University of Medicine, Bari, Italy

<sup>7</sup> Univ. of Bordeaux, ISPED, LESIM, F-33060, Bordeaux, France

**Abstract.** Rates of healthcare databases offers great potential for producing timely epidemiological evidence reflecting 'real world' circumstances. The advantage of such databases is maximized when data from different countries and organizations are pooled. Several ongoing initiatives are facing the technical and scientific challenges of pooling data from substantially diverse sources. In this workshop some notable experiences from USA and Europe will be presented and compared.

**Keywords.** data reuse, observational studies, pharmacoepidemiology, health services research.

### Introduction

Several ongoing initiatives are aimed at exploiting existing administrative claims and electronic medical record databases from different institutions/countries to automatically produce datasets for the conduct of epidemiological studies. Although all the healthcare databases involved in these projects share a basic semantic similarity, challenges arise in intelligent, semantic interoperability from the heterogeneity of the

\* Corresponding Author.

# Indice

Introduzione

Obbiettivi e struttura del progetto MATRICE

Software

Gli studi di validazione

TheMatrix oltre MATRICE

Discussione

# Discussione





Grazie per l'attenzione!