

"TheShinISS": un applicativo open-source per la conduzione di analisi distribuite in studi di farmacoepidemiologia di tipo multi-database

Marco Massari, Stefania Spila Alegiani, Roberto Da Cas, Francesca Menniti Ippolito

Centro Nazionale per la Ricerca e la Valutazione Preclinica e Clinica dei Farmaci, Istituto Superiore di Sanità, Roma

SUMMARY

"TheShinISS": an open-source tool for conducting distributed analyses within pharmacoepidemiological multi-database studies

Introduction

Healthcare databases represent useful source of data for conducting pharmacoepidemiological studies on drug and vaccine utilization, efficacy and safety. When the studies include different geographical areas, it is appropriate to use a multi-database study approach, based on a Common Data Model (CDM), to conduct locally distributed analyses. This article describes the experience of the Istituto Superiore di Sanità (ISS) - Italian National Institute of Health in planning and implementing a statistical-informatics tool for conducting multi-database studies based on a Common Data Model (CDM).

Materials and methods

"TheShinISS" is an R open-source tool developed by ISS for conducting distributed analyses within the main epidemiological multi-database study designs: descriptive, cohort, case-control, case-cohort, self-controlled-case-series. "TheShinISS" allows elaborating and processing health archives at local level, performing data quality control, matching/sampling, record-linkage, and finally creating the anonymized dataset for the centralized data analyses.

Results

"TheShinISS" was used in pharmacoepidemiological studies conducted within the ITA-COVID19 network, which includes ISS, universities and regions. The network was set up during the pandemic emergency to promote the conduction of observational studies to provide prompt evidence on the role of drugs and vaccines on the prognosis of patients with COVID-19.

Conclusions

"TheShinISS" proved to be easy to use and adaptable to different research aims and designs. Through the reproducibility of every face of the analyses, it has provided satisfying quality control in the creation of analytical datasets. These features made it possible the inclusion of different regions/provinces in the ITA-COVID19 network, and above all reduced the time necessary for the conduction of studies.

Key words: pharmacoepidemiology, electronic health records; multi-database study

marco.massari@iss.it

Introduzione

Gli archivi sanitari rappresentano fonti di dati di grande utilità per la conduzione di studi post-marketing sull'uso, la sicurezza e l'efficacia dei farmaci e dei vaccini, in particolare nelle popolazioni che non sono incluse nei trial pre-registrativi, come la popolazione pediatrica e anziana. Inoltre, i flussi di dati sanitari correnti possono fornire informazioni utili prima della commercializzazione dei farmaci, come ad esempio, la stima del numero di pazienti che potrebbero beneficiare di trattamenti innovativi e ad alto costo. Quindi, attraverso la conduzione di studi di farmacoepidemiologia sia in fase pre- che post-registrativa, tali fonti informative sono in grado di generare evidenze per supportare le agenzie regolatorie nell'elaborazione e nell'implementazione di provvedimenti per rispondere tempestivamente a problematiche rilevanti per la sanità pubblica. Nel caso di studi

che analizzano esposizioni o esiti poco frequenti, oppure nel caso di studi rivolti a sottogruppi di popolazione, l'utilizzo di archivi sanitari correnti riferiti a una sola area geografica anche di grandi dimensioni, può non essere sufficiente per il corretto dimensionamento dello studio. In tali situazioni, per garantire una potenza statistica adeguata, è necessario condurre studi multi-database combinando banche dati di più aree.

Si definisce multi-database "uno studio che utilizza almeno due banche dati sanitarie, che non sono collegate tra loro a livello individuale, o perché fanno riferimento a popolazioni differenti, o perché, sebbene riguardino la stessa popolazione, le normative locali impediscono il record linkage" (1).

Nell'articolo di Gini e colleghi (1) e nella linea guida dell'European Network of Centers for Pharmacovigilance (ENCePP) sugli standard metodologici in farmacoepidemiologia (2) sono

state individuate quattro differenti strategie per la conduzione di studi multi-database all'interno di un network di diverse aree geografiche (ad esempio, Regioni o singole aziende sanitarie). Tali strategie sono state schematizzate sulla base delle modalità di creazione del dataset analitico e di conduzione delle analisi in:

a) **Analisi locali**: i dati sono estratti e analizzati esclusivamente a livello locale attraverso script di analisi sviluppati all'interno di ciascun Centro della rete. Un esempio di questa strategia di conduzione degli studi è rappresentato dal Canadian Network for Observational Drug Effect Studies (3);

b) **Condivisione degli archivi sanitari (dati raw)**: i record e le variabili di interesse per lo studio vengono estratti dagli archivi a livello locale e trasferiti senza alcun tipo di analisi a un Centro di riferimento che effettua l'aggregazione dei flussi e l'analisi *pooled* dei dati; un esempio europeo è rappresentato dal Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria (4);

c) **Uso di un Common Data Model "studio-specifico"**: i dati raw e le variabili dei flussi sanitari, di interesse per lo studio, vengono inizialmente estratti e pre-processati localmente al fine di creare una struttura dei dati comune a tutti i centri che afferiscono alla rete (Common Data Model - CDM). Successivamente, sempre a livello locale, i dati strutturati secondo il CDM vengono elaborati attraverso uno script di analisi comune inviato a tutti i centri partecipanti dal Centro di riferimento (analisi distribuita). L'output viene poi restituito al Centro di riferimento ed è in genere costituito da un dataset analitico, o in alcuni casi, da risultati aggregati. Diversi progetti a livello italiano ed europeo hanno applicato questa strategia (5-11);

d) **Uso di un CDM "generale"**: i dati raw e le variabili dei flussi sanitari vengono estratti e caricati a livello locale in un CDM permanente e, indipendentemente da uno specifico studio, vengono aggiornati periodicamente. Ogni volta che un nuovo studio viene autorizzato e condiviso tra i centri partecipanti, si procede in modo del tutto analogo alla strategia c) attraverso un sistema di analisi distribuite. Un esempio di applicazione di questa strategia è rappresentato da FDA's Sentinel Initiative (12).

Per poter essere realizzate, le strategie c) e d) richiedono la progettazione e l'impiego di specifici strumenti statistico-informatici al fine di implementare il sistema di analisi distribuite nella rete rispettando al contempo la normativa sulla privacy. Alcune applicazioni di questo tipo sono rappresentate in Italia dal software "TheMatrix" (13) e a livello europeo dal software Jerboa© (6).

In questo lavoro viene presentata l'esperienza dell'Istituto Superiore di Sanità (ISS) nella progettazione e realizzazione di un applicativo statistico-informatico che permette la conduzione di studi multi-database basati su analisi distribuite attraverso l'impiego di un CDM "studio-specifico" (strategia c).

Materiali e metodi

"TheShinISS" (Figura 1) è un progetto open-source sviluppato con il software statistico R (14), concepito e realizzato dai ricercatori del Reparto di Farmacoepidemiologia e Farmacosorveglianza del Centro Nazionale per la Ricerca e la Valutazione Preclinica e Clinica dei Farmaci (CNRVF) dell'ISS per la conduzione di analisi distribuite applicate ai principali disegni di studi epidemiologici multi-database (descrittivi, coorte, caso-controllo, caso-coorte, self-controlled-case-series). Il nome dell'applicativo proviene dalla libreria Shiny, del software R, che permette di creare front-end interattivi (15). Tra gli aspetti caratterizzanti di "TheShinISS" vi sono: facilità di utilizzo da parte dell'utente, grazie a un'interfaccia grafica che non richiede la conoscenza di R; controlli di qualità sui flussi prima che questi vengano elaborati; reportistica flessibile, dettagliata e riproducibile grazie al linguaggio R Markdown (16,17) ed efficienza nell'elaborazione dei dati grazie all'insieme di librerie note con il nome di Tidyverse (18). Infine, il sistema risulta particolarmente stabile in quanto indipendente dall'aggiornamento delle librerie di R grazie alla gestione delle dipendenze *renv* (19) che rende il progetto "TheShinISS" portabile e completamente riproducibile.

Nel descrivere le fasi operative di utilizzo di "TheShinISS", applicato a uno studio multi-database basato su CDM, si farà riferimento come esempio a uno studio caso-controllo multi-regionale dove i casi sono rappresentati dai soggetti positivi al test SARS-CoV-2 identificati dai registri regionali di sorveglianza (registro COVID-19) e i controlli sono rappresentati dagli assistibili riferiti alla popolazione regionale di provenienza dei casi (anagrafica assistibili).

Le Regioni o le singole aziende sanitarie partecipanti allo studio (Centri), dopo aver condiviso il protocollo, ricevono l'applicativo "TheShinISS" insieme alla Procedura Operativa Standard (SOP) che descrive nel dettaglio le fasi di seguito riportate.

Fase 1 - Estrazione e predisposizione a livello locale di un CDM rappresentato dal registro dei soggetti positivi SARS-CoV-2 e dall'anagrafica degli assistibili.

TheShinISS (Distributed Analyses)

TheShinISS (ver. 1.3 del 26-05-2020)
SOP di Riferimento: ITA-COVID19-CQ-HCQ (ver. 1.0 del 26-05-2020)
Associazione fra Cloroquina/drossicloroquina e rischio di ospedalizzazione da COVID-19

Requisiti di sistema

Computer 64 bit (necessari 16 GB di RAM per Regioni con popolazione > 5 ML di abitanti)
R versione 3.6.x
Rtools versione 3.5 (<https://cran.r-project.org/bin/windows/Rtools/history.html>)
Rstudio versione > 0.98.945 (consigliata l'ultima)

Aggiornamenti Dashboard

1.3 (26/02/2020)

- 1a Release per studi multi-database caso-controllo

Aggiornamenti Script

1.0.2 (03/06/2020)

- Correzione data su quality check: dataric=="2020-02-21"
- Inserimento filtro di sicurezza [18-120] per la variabile eta nel flusso Anagrafica
- Correzione bug su quality check dataguar che restituisce: ERROR: no applicable method for "filter_" applied to an object of class "NULL"

1.0.1 (28/05/2020)

- Correzione bug nella generazione dell'output 'id_mtc.csv': selezione dei casi e dei controlli
- Controllo plausibilità (dataric <= dataguar) solo se dataguar è non missing

1.0.0 (26/05/2020)

- 1a release specifica per il progetto: script per il controllo di qualità e la generazione dataset analitico

Riferimenti per 'TheShinISS'

Marco Massari	marco.massari@iss.it	Skype: massarimar
Stefania Spila Alegiani	stefania.spila@iss.it	Skype: ssa-iss

Figura 1 - Front-end dell'applicativo "TheShinISS"

Fase 2 - Esecuzione dell'applicativo sul CDM locale, per l'identificazione della popolazione in studio sulla base del disegno epidemiologico scelto (in questo caso l'applicativo esegue un algoritmo di matching individuale, per esempio per età e sesso, al fine di estrarre casi e controlli con un rapporto 1:n).

Fase 3 - Estrazione e predisposizione a livello locale dei flussi correnti sulla base del CDM nella popolazione in studio selezionata. Esempi di flussi possono riguardare le prescrizioni farmaceutiche, l'anagrafe vaccinale, le schede di dimissione ospedaliera, le esenzioni o altri tipi di archivi sanitari locali purché collegabili tra loro attraverso un identificativo anonimo univoco.

Fase 4 - Esecuzione a livello locale dell'applicativo sul CDM per effettuare: il linkage della popolazione in studio con i flussi sanitari correnti, l'aggregazione e la creazione del minimo insieme di variabili di esposizione e di covariate di interesse specifico per lo studio che andranno a costituire il dataset analitico anonimizzato (Figura 2).

Tutti i dataset locali vengono successivamente resi disponibili al Centro di riferimento che li combina (merge) per la creazione del dataset di analisi finale. Le fasi 1 e 3 sono di pertinenza dei Centri e richiedono un impegno contenuto di data-management limitatamente all'estrazione dei flussi sanitari e alla predisposizione di questi secondo un CDM. Infatti, nella maggior parte dei casi, ai Centri viene richiesto di trasformare gli archivi originali in file in formato testo (.csv) con una denominazione comune delle variabili e di ricodificare alcune di esse secondo modalità e formati condivisi.

Tutta l'attività di elaborazione, processamento dei flussi, controllo di qualità, matching/campionamento, record-linkage, aggregazione, e per finire creazione del minimo set di variabili necessarie per lo studio che costituirà il dataset di analisi anonimizzato, viene invece eseguita dall'applicativo "TheShinISS" (Fasi 2 e 4).

Risultati

L'applicativo "TheShinISS" è stato utilizzato in studi di farmacoepidemiologia multi-database condotti all'interno della rete ITA-

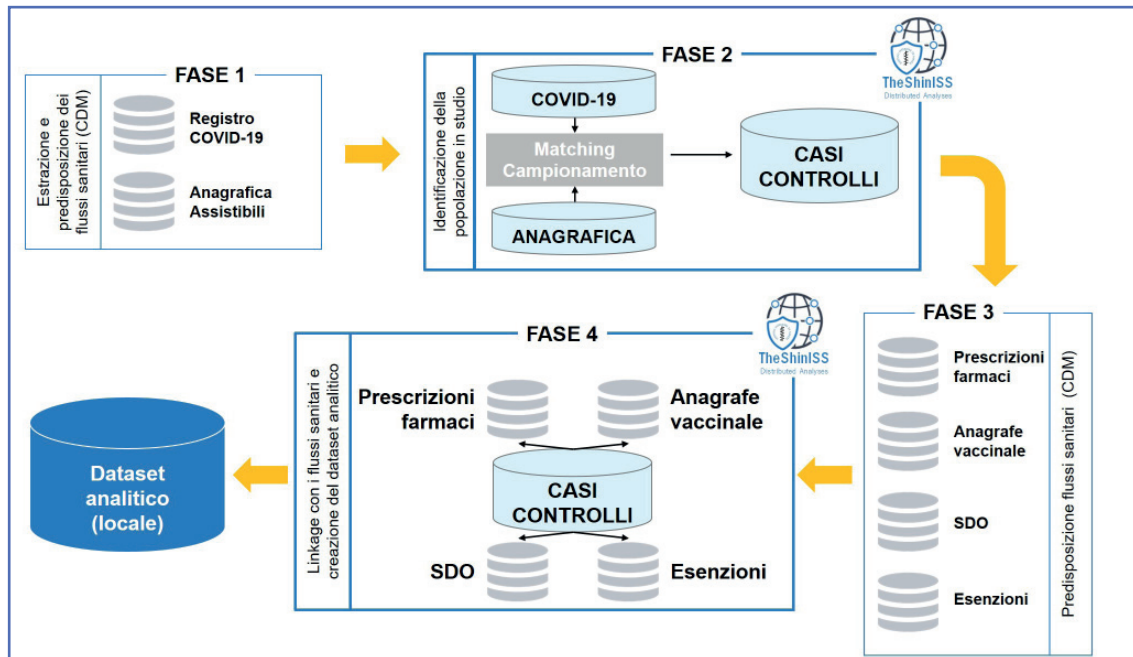


Figura 2 - Schema delle fasi operative di analisi distribuite di un progetto attraverso “TheShinISS”. CDM: Common Data Model; SDO: schede di dimissione ospedaliera

COVID19. Tale rete, coordinata dal Reparto di Farmacoepidemiologia e Farmacosorveglianza del CNRVF, in collaborazione con il Dipartimento di Malattie Infettive dell’ISS, si è costituita tra alcune Regioni, aziende sanitarie e università durante l’emergenza pandemica. La rete, nata con finalità di ricerca, ha l’obiettivo di favorire la rapida conduzione di studi osservazionali multi-database in corso di pandemia per contribuire a fornire evidenze in tempi brevi sul ruolo di farmaci o vaccini sull’insorgenza di infezione da SARS-CoV-2 e sul rischio di ospedalizzazione o morte in pazienti affetti da COVID-19 (20-27).

Il primo studio condotto attraverso la rete ITA-COVID19 ha riguardato il ruolo delle terapie farmacologiche con ACE inibitori (ACEIs) e/o sartani (ARBs) sulla prognosi dei pazienti con COVID-19 (28).

Questa prima esperienza ha dimostrato l’utilità dell’applicativo “TheShinISS” come strumento statistico-informatico nella conduzione di uno studio multi-database basato su CDM e ha permesso di valorizzare il modello organizzativo di rete multi-regionale, garantendo risposte tempestive (lo studio è stato condotto in meno di 2 mesi), anche a supporto di decisioni regolatorie. In tal senso, al fine di fornire una stima della sopravvivenza, da utilizzare come comparatore in studi di fase II, è stato condotto uno studio di mortalità nella coorte di pazienti ricoverati con COVID-19 (29).

Sono inoltre in fase di conclusione uno studio caso-controllo nested in una coorte di utilizzatori di farmaci antireumatici “conventional Synthetic Disease Modifying Antirheumatic Drugs” (cDMARDs), che aveva l’obiettivo di valutare l’utilizzo precedente di idrossiclorochina o cloroquina sulla prognosi in pazienti COVID-19 e uno studio di coorte, sull’associazione tra la vaccinazione antinfluenzale durante la stagione 2019-2020 e il rischio di ospedalizzazione, ricovero nelle terapie intensive e mortalità in pazienti con COVID-19.

Discussione

In Italia, i primi studi di farmacoepidemiologia che hanno utilizzato archivi sanitari correnti, collegati tra loro attraverso metodi di record-linkage, sono stati condotti dall’ISS negli anni ‘90, nell’ambito del progetto Videofar (30-32). Tale progetto aveva l’obiettivo di definire modelli di analisi degli archivi di prescrizione farmaceutica e lo sviluppo di indicatori per la conduzione di studi sull’uso e sulla sicurezza dei farmaci. Successivamente, attraverso il sistema Babele, sono state riconosciute le potenzialità derivanti dal processo di integrazione di più fonti di dati correnti ed è stata quindi formalizzata la metodologia per gli studi di farmacoepidemiologia basati sull’utilizzo degli archivi elettronici sanitari (33).

Nel corso degli anni il Reparto di Farmacoepidemiologia e Farmacosorveglianza ha condotto studi di farmacoepidemiologia basati

sulla condivisione di dati raw provenienti da archivi sanitari (34-36) ma, in seguito alla nuova normativa sulla privacy (37), gli studi effettuati con tale metodologia sono diventati di difficile conduzione.

L'applicativo "TheShinISS" è stato ideato e sviluppato come progetto pilota nell'ambito delle attività formative dei corsi ISS sull'analisi delle prescrizioni farmaceutiche come esempio di analisi riproducibile su archivi sanitari correnti, e successivamente portato a completa realizzazione, come strumento software per la conduzione di analisi distribuite nell'ambito di studi multi-database su CDM durante l'emergenza COVID-19 a supporto della appena costituita rete ITA-COVID19.

Durante l'esperienza della rete ITA-COVID19 (28, 29), l'applicativo "TheShinISS" ha messo in evidenza i numerosi vantaggi della conduzione degli studi multi-database basati su CDM studio-specifico. Infatti, mentre la tempistica per l'estrazione e la predisposizione dei flussi secondo un CDM (Fasi 1 e 3) dipende dalle risorse locali dei Centri e dalla complessità del processo di armonizzazione dei differenti archivi locali, la tempistica per l'elaborazione degli archivi e per la generazione del dataset analitico, attraverso "TheShinISS" (Fasi 2 e 4), viene ridotta a pochi minuti di calcolo per regioni di grandi dimensioni, garantendo al contempo il massimo livello di trasparenza.

Ovviamente se la strategia basata su analisi distribuite rappresenta un evidente risparmio di risorse a livello locale, l'adattamento dell'applicativo a un nuovo progetto richiede, in alcuni casi, un investimento di tempo non trascurabile, specialmente per la fase di rilascio dell'applicativo dove è prevista un'approfondita fase di test su dati simulati. Tuttavia questo processo può avvenire nelle fasi precoci dello studio, ad esempio mentre viene finalizzata la SOP, riutilizzando dove possibile intere sezioni di codice R, come quello che effettua controlli di qualità o genera il set di covariate. Inoltre, in base alla filosofia di programmazione agile, il codice può essere migliorato e reso più efficiente man mano che viene adattato a nuovi progetti di studio. Infine la centralizzazione degli script riduce la possibilità di generare errori quando le analisi vengono condotte a livello locale con inevitabile duplicazione degli script o la non aderenza alle specifiche della SOP (strategia a).

Un elemento essenziale dell'applicativo riguarda il controllo di qualità degli archivi che viene effettuato a ogni esecuzione sull'aderenza alle specifiche del CDM creato da ciascun Centro e sugli archivi che costituiscono il modello di dati comune. Tale operazione è importante poiché il Centro di riferimento per l'analisi centralizzata, non avendo accesso agli archivi sanitari di origine,

non ha la possibilità di verificarne la qualità dei dati. L'applicativo, al termine del controllo di qualità, fornisce una reportistica dettagliata che viene inviata al Centro di riferimento insieme al dataset analitico.

Infine, un ulteriore aspetto saliente riguarda il trattamento dei dati. È noto che i database amministrativi gestiti e condivisi come dati raw (strategia b), per quanto anonimizzati, contengono dati personali sensibili costituiti da più record per ogni soggetto (ad esempio, l'archivio delle prescrizioni farmaceutiche) e questo aumenta la probabilità di individuare i soggetti in studio. Al contrario, la strategia basata su CDM a cui fa riferimento l'applicativo "TheShinISS", attraverso l'elaborazione locale dei dati raw e la condivisione del solo dataset analitico anonimizzato non rende possibile risalire all'identificazione dei soggetti (37). Infatti, il dataset analitico contiene il minimo set di variabili necessarie per il quesito di ricerca, il più possibile aggregate, e opportunamente trasformate (ad esempio, trasformazione delle date di nascita in classi di età, delle date degli eventi in settimane o mesi di calendario o in intervalli di tempo all'evento).

Conclusioni

"TheShinISS" si è dimostrato di facile uso a livello locale e flessibile nell'adattarsi ai differenti quesiti di ricerca. Attraverso la riproducibilità di tutte le fasi di analisi, secondo un flow-chart predefinito, l'applicativo ha garantito un controllo di qualità adeguato nell'elaborazione dei flussi e nella creazione dei dataset analitici. Tali caratteristiche hanno reso possibile la partecipazione di più Regioni/ASL nella rete ITA-COVID19, e soprattutto la riduzione dei tempi di conduzione degli studi. Al momento, il processo di documentazione del codice sorgente è ancora a una fase iniziale e l'applicativo risulta ancora non facilmente riadattabile da un programmatore esterno; questa limitazione sarà in futuro superata rendendo disponibile una documentazione dettagliata basata su template di codice commentato.

L'esperienza maturata finora ha permesso e permetterà di adattare l'applicativo "TheShinISS" ad altri studi sia all'interno della rete ITA-COVID19 che in altri ambiti di ricerca, come il progetto VALORE "VALutazione post-marketing del profilo beneficio-rischio dei farmaci biologici Originator e biosimilari in area dermatologica, reumatologica, gastroenterologica e oncematologica tramite la costituzione di un network unico multiregionale per l'analisi integrata di dati provenienti da banche dati sanitarie, sorveglianze attive e Registri clinici"; il progetto CESIT "Valutazione

comparativa di efficacia e sicurezza dei farmaci immunosoppressori nei pazienti trapiantati” (progetti multiregionali di farmacovigilanza AIFA); i progetti TheShinISS-VAX sulla sicurezza dei vaccini antinfluenzali e dei vaccini anti-COVID-19.

Ringraziamenti

Un particolare ringraziamento va a Gianluca Trifirò (Università di Verona) e a Valeria Belleudi (Dipartimento di Epidemiologia del Servizio Sanitario Regionale del Lazio) e a tutti i colleghi (Andrea Fontana, Valentina Ientile, Ylenia Ingrassiotta, Valentina Isgro, Luca L'Abbate, Gabriele Puglisi, Janet Sultana - Università di Messina; Marco Finocchietti, Francesca Poggi, Alessandro Rosa - Dipartimento di Epidemiologia del Servizio Sanitario Regionale del Lazio) che hanno contribuito a rendere più agile l'applicativo adattandolo ai nuovi progetti e rendendo più efficiente parte del codice.

Citare come segue:

Massari M, Spila Alegiani S, Da Cas R, Menniti Ippolito F. TheShinISS: un applicativo open-source per la conduzione di analisi distribuite in studi di farmacoepidemiologia di tipo multi-database. *Boll Epidemiol Naz* 2020; 1(2):39-45.

Conflitti di interesse dichiarati: nessuno.

Finanziamenti: nessuno.

Authorship: tutti gli autori hanno contribuito in modo significativo alla realizzazione di questo studio nella forma sottomessa.

Riferimenti bibliografici

- Gini R, Sturkenboom MCJ, Sultana J, Cave A, Landi A, Pacurariu A, et al. Different strategies to execute multi-database studies for medicines surveillance in real-world setting: a reflection on the European model. *Clin Pharmacol Ther* 2020;108(2):228-35; doi: 10.1002/cpt.1833
- The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on methodological standards in pharmacoepidemiology. 2020. http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml ultimo accesso 2/12/2020.
- Suissa S, Henry D, Caetano P, Dormuth CR, Ernst P, Hemmelgarn B, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med* 2012;6(4):e134-40; PMID: 23687528; PMC3654509
- Maciá Martínez MA, Gil M, Huerta C, Martín Merino E, Álvarez A, Bryant V, et al. Base de datos para la investigación farmacoepidemiológica en atención primaria (BIFAP): a data resource for pharmacoepidemiology in Spain. *Pharmacoepidemiol Drug Saf* 2020;29(10):1236-45; doi: 10.1002/pds.5006
- Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *J Intern Med* 2014;275(6):551-61; doi: 10.1111/joim.12159
- Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf* 2011;20(1):1-11; doi:10.1002/pds.2053
- Arfè A, Scotti L, Varas-Lorenzo C, Nicotra F, Zambon A, Kollhorst B, et al. Non-steroidal anti-inflammatory drugs and risk of heart failure in four European countries: nested case-control study. *BMJ* 2016;354:i4857; doi: 10.1136/bmj.i4857
- Trifirò G, de Ridder M, Sultana J, Oteri A, Rijnbeek P, Pecchioli S, et al. Use of azithromycin and risk of ventricular arrhythmia. *CMAJ* 2017;189(15):E560-8; doi: 10.1503/cmaj.160355
- Dieleman J, Romio S, Johansen K, Weibel D, Bonhoeffer J, Sturkenboom M. Guillain-Barre syndrome and adjuvanted pandemic influenza A (H1N1) 2009 vaccine: multinational case-control study in Europe. *BMJ* 2011;343:d3908; doi: 10.1136/bmj.d3908
- Becker BFH, Avillach P, Romio S, van Mulligen EM, Weibel D, Sturkenboom MCJM, et al. CodeMapper: semiautomatic coding of case definitions. A contribution from the ADVANCE project. *Pharmacoepidemiol Drug Saf* 2017;26(8):998-1005; doi: 10.1002/pds.4245
- Roberto G, Leal I, Sattar N, Loomis AK, Avillach P, Egger P, et al. Identifying cases of type 2 diabetes in heterogeneous data sources: strategy from the EMIF Project. *PLoS One* 2016;11(8):e0160648; doi: 10.1371/journal.pone.0160648
- Platt R, Brown JS, Robb M, McClellan M, Ball R, Nguyen MD, et al. The FDA sentinel initiative - An evolving national resource. *N Engl J Med* 2018;379(22):2091-3; doi: 10.1056/NEJMp1809643
- Gini R, Schuemie M, Brown J, Ryan P, Vacchi E, Coppola M, et al. Data extraction and management in networks of observational health care databases for scientific research: a comparison among EU-ADR, OMOP, Mini-Sentinel and MATRICE Strategies. *EGEMS (Wash DC)* 2016;4(1):1189; doi: 10.13063/2327-9214.1189
- R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna; 2013. <http://www.R-project.org/>; ultimo accesso 2/12/2020.
- Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Shiny: Web Application Framework for R. 2020. <https://CRAN.R-project.org/package=shiny> ultimo accesso 2/12/2020.
- Xie Y, Dervieux C, Riederer E. R Markdown Cookbook. Chapman and Hall/CRC. 2020. <https://bookdown.org/yihui/rmarkdown-cookbook> ultimo accesso 2/12/2020.
- Xie Y, Allaire JJ, Golemund G. R Markdown: The definitive guide. Chapman and Hall/CRC. 2020. <https://bookdown.org/yihui/rmarkdown> ultimo accesso 2/12/2020.

18. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. *J Open Source Softw* 2019;4(43):1686; doi: 10.21105/joss.01686
19. Ushey K. Renv: Project Environments. 2020. <https://rstudio.github.io/renv/> ultimo accesso 2/12/2020.
20. Gurwitz D. Angiotensin receptor blockers as tentative SARS-CoV-2 therapeutics. *Drug Dev Res* 2020;81(5):537-40; doi: 10.1002/ddr.21656
21. Zheng YY, Ma YT, Zhang JY, Xie X. COVID-19 and the cardiovascular system. *Nat Rev Cardiol* 2020;17(5):259-60; doi: 10.1038/s41569-020-0360-5
22. Watkins J. Preventing a COVID-19 pandemic. *BMJ* 2020;368:m810; doi: 10.1136/bmj.m810
23. Palacios G, Hornig M, Cisterna D, Savji N, Bussetti AV, Kapoor V, et al. Streptococcus pneumoniae coinfection is correlated with the severity of H1N1 pandemic influenza. *PLoS One* 2009;4(12):e8540; doi: 10.1371/journal.pone.0008540
24. Leentjens J, Kox M, Stokman R, Gerretsen J, Diavatopoulos DA, van Crevel R, et al. BCG Vaccination enhances the immunogenicity of subsequent influenza vaccination in healthy volunteers: a randomized, placebo-controlled pilot study. *J Infect Dis* 2015;212(12):1930-8; doi: 10.1093/infdis/jiv332
25. Gao J, Tian Z, Yang X. Breakthrough: chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. *Biosci Trends* 2020;14(1):72-3; doi: 10.5582/bst.2020.01047
26. Touret F, de Lamballerie X. Of chloroquine and COVID-19. *Antiviral Res* 2020;177:104762; doi: 10.1016/j.antiviral.2020.104762
27. Cortegiani A, Ingoglia G, Ippolito M, Giarratano A, Einav S. A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. *J Crit Care* 2020;57:279-83; doi: 10.1016/j.jcrc.2020.03.005
28. Trifirò G, Massari M, Da Cas R, Menniti Ippolito F, Sultana J, Crisafulli S, et al. Renin-angiotensin-aldosterone system inhibitors and risk of death in patients hospitalised with COVID-19: a retrospective Italian cohort study of 43,000 patients. *Drug Saf* 2020;43(12):1297-308; doi: 10.1007/s40264-020-00994-5
29. Ferroni E, Giorgi Rossi P, Spila Alegiani S, Trifò G, Pitter G, Leoni O, et al. Survival of hospitalized COVID-19 patients in Northern Italy: a population-based cohort study by the ITA-COVID19 Network. *Clin Epidemiol* 2020;12:1337-46; doi: <https://doi.org/10.2147/CLEP.S271763>
30. Caffari B, Raschetti R. The logical structure of the VIDEOFAR drug data base. *Ann Ist Super Sanita* 1991;27(2):195-200; PMID: 1755570
31. Maggini M, Menniti Ippolito F, Spila Alegiani S, Traversa G, Fortini M. Drug utilization studies within the VIDEOFAR project. *Ann Ist Super Sanita* 1991;27(2):201-6; PMID: 1755571
32. Raschetti R. VIDEOFAR: an epidemiologic tool for drug utilization studies. *Ann Ist Super Sanita* 1991;27(2):191-3; PMID: 1755569
33. Caranci N, Fano V, Gini R, Maggini M, Raschetti R, Simonato L. A laboratory to overcome the babel of the electronic health archives. *Epidemiol Prev* 2012;36(5):234-5; PMID: 23139108
34. Trotta F, Spila Alegiani S, Da Cas R, Rajevic M, Conti V, Venegoni M, et al. Cardiovascular safety of tiotropium respimat vs handihaler in the routine clinical practice: a population-based cohort study. *PLoS One* 2017;12(4):e0176276; doi: 10.1371/journal.pone.0176276
35. Spila Alegiani S, Trotta F, Da Cas R, Rossi M, Venegoni M, Traversa G. Comparative effectiveness of two tiotropium formulations: a retrospective cohort study. *COPD* 2018;15(5):418-23; doi: 10.1080/15412555.2018.1554032
36. Da Cas R, Nigro A, Terrazzino S, Sances G, Viana M, Tassorelli C, et al. Triptan use in Italy: insights from administrative databases. *Cephalalgia* 2015;35(7):619-26; doi: 10.1177/0333102414550419
37. Regolamento 2016/679-GDR del Parlamento Europeo e del Consiglio del 27 aprile 2016.