OHDSI Literature 2019-2020

Vocabulary:
Mapping local codes

Comparative Analysis and Evaluation of State-of-the-Art Medication Mapping Tools to Transform a Local Medication Terminology to RxNorm

AMIA Jt Summits Transl Sci Proc. 2020 May 30;2020:126-135. eCollection 2020.

Authors

Lena Davidson, Mary Regina Boland

Abstract

Mapping local terminologies to standardized terminologies facilitates secondary use of electronic health records (EHR). Penn Medicine comprises multiple hospitals and facilities within the Philadelphia Metropolitan area providing services from primary to quaternary care. Our Penn Medicine (PennMed) data include medications collected during both inpatient and outpatient encounters at multiple facilities. Our goal was to map 941,198 unique medication terms to RxNorm, a standardized drug nomenclature from the National Library of Medicine (NLM). We chose three popular tools for mapping: NLM's RxMix and RxNav-in-a-Box, OHDSI's Usagi and Mayo Clinic's MedXN. We manually reviewed 400 mappings obtained from each tool and evaluated their performance for drug name, strength, form, and route. RxMix performed the best

Comparison of the cohort selection performance of Australian Medicines Terminology to Anatomical Therapeutic Chemical mappings

J Am Med Inform Assoc. 2019 Nov 1;26(11):1237-1246. doi: 10.1093/jamia/ocz143.

Authors

Guan N Guo, Jitendra Jonnagaddala, Sanjay Farshid, Vojtech Huser, Christian Reich, Siaw-Teng Liaw

Abstract

Objective: Electronic health records are increasingly utilized for observational and clinical research. Identification of cohorts using electronic health records is an important step in this process. Previous studies largely focused on the methods of cohort selection, but there is little evidence on the impact of underlying vocabularies and mappings between vocabularies used for cohort selection. We aim to compare the cohort selection performance using Australian Medicines Terminology to Anatomical Therapeutic Chemical (ATC) mappings from 2 different sources. These mappings were taken from the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) and the Pharmaceutical Benefits Scheme (PBS) schedule.

Feasibility of Mapping Austrian Health Claims Data to the OMOP Common Data Model

J Med Syst. 2019 Sep 7;43(10):314. doi: 10.1007/s10916-019-1436-9.

Authors

Andrea Haberson, Christoph Rinner, Alexander Schöberl, Walter Gall

Abstract

The Main Association of Austrian Social Security Institutions collects pseudonymized claims data from Austrian social security institutions and information about hospital stays in a database for research purposes. For new studies the same data are repeatedly reprocessed and it is difficult to compare different study results even though the data is already preprocessed and prepared in a proprietary data model. Based on a study on adverse drug events in relation to inappropriate medication in geriatric patients the suitability of the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) is analyzed and data is transformed into the OMOP CDM. 1,023 (99.7%) of drug codes and 3,812 (99.2%) of diagnoses codes coincide with the OMOP vocabularies. The biggest obstacles are missing mappings for the Local Vocabularies like the Austrian pharmaceutical registration numbers and the Socio-Economic Index to the OMOP vocabularies. OMOP CDM is a promising approach for the standardization of Austrian claims data.

Standardizing Austrians Claims Data Using the OMOP Common Data Model: A Feasibility Study

Stud Health Technol Inform. 2019;258:151-152.

Authors

Andrea Haberson, Christoph Rinner, Walter Gall

Abstract

The suitability of the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) for Austrian pseudonymized claims data from social security institutions and information about hospital stays is evaluated. 1,023 (99.7%) of ATC codes and 3,695 (98.6%) of ICD10 codes coincide with the OMOP vocabulary. Mappings for the local vocabularies like the Austrian pharmaceutical registration numbers, the Socio-Economic Index and professional groups, to the OMOP vocabulary do not exist. A standardization with the OMOP CDM is possible, however an initial, not negligible effort is required to adapt and incorporate the vocabulary.

Transforming French Electronic Health Records into the Observational Medical Outcome Partnership's Common Data Model: A Feasibility Study

Appl Clin Inform. 2020 Jan;11(1):13-22. doi: 10.1055/s-0039-3402754. Epub 2020 Jan 8.

Authors

Antoine Lamer, Nicolas Depas, Matthieu Doutreligne, Adrien Parrot, David Verloop, Marguerite-Marie Defebvre, Grégoire Ficheur, Emmanuel Chazard, Jean-Baptiste Beuscart

Abstract

Background: Common data models (CDMs) enable data to be standardized, and facilitate data exchange, sharing, and storage, particularly when the data have been collected via distinct, heterogeneous systems. Moreover, CDMs provide tools for data quality assessment, integration into models, visualization, and analysis. The observational medical outcome partnership (OMOP) provides a CDM for organizing and standardizing databases. Common data models not only facilitate data integration but also (and especially for the OMOP model) extends the range of available statistical analyses.

Objective: This study aimed to evaluate the feasibility of implementing French national electronic

Common data model: Creating local databases

Data Integration into OMOP CDM for Heterogeneous Clinical Data Collections via HL7 FHIR Bundles and XSLT

Stud Health Technol Inform. 2020 Jun 16;270:138-142. doi: 10.3233/SHTI200138.

Authors

Patrick Fischer, Mark R Stöhr, Henning Gall, Achim Michel-Backofen, Raphael W Majeed

Abstract

Data integration is an important task in medical informatics and highly impacts the gain out of existing health information data. These tasks are using implemented as extract transform and load processes. By introducing HL7 FHIR as an intermediate format, our aim was to integrate heterogeneous data from a German pulmonary hypertension registry into an OMOP Common Data Model. First, domain knowledge experts defined a common parameter set, which was subsequently mapped to standardized terminologies like LOINC or SNOMED-CT. Data was extracted as HL7 FHIR Bundle to be transformed to OMOP CDM by using XSLT. We successfully transformed the majority of data elements to the OMOP CDM in a feasible time.

Specification and Distribution of Vocabularies Among Consortial Partners

Stud Health Technol Inform. 2020 Jun 16;270:1393-1394. doi: 10.3233/SHTI200458.

Authors

Mirko Gruhl, Ines Reinecke, Martin Sedlmayr

Abstract

Due to the variety of different software systems and disparate observational databases, the need for a uniform data representation rises. Common data models (CDM) support the harmonisation of data. A powerful but compact software setup and a minimum vocabulary set has been composed via Docker to facilitate analysis of data across ten university hospitals. The presented approach also creates the possibility to use a concise database which is easy to deploy.

Real-world data in Saudi Arabia: Current situation and challenges for regulatory decision-making

Pharmacoepidemiol Drug Saf. 2020 May 27. doi: 10.1002/pds.5025. Online ahead of print.

Authors

Fatemah A Alnofal, Adel A Alrwisan, Thamir M Alshammari

Abstract

Purpose: To present the process of establishing a pharmacoepidemiological database in Saudi Arabia, challenges and models used.

Methods: The database establishment has started in 2017 by piloting the conversion of electronic health records of one hospital to the Observational Health Data Sciences and Informatics (OHDSI), Observational Medical Outcomes Partnership's Common Data Model (OMOP).

Results: During the pilot phase we have faced several challenges such as limited contribution in providing data by local medical institution due to uncertainty about data governance, diversity of systems used by hospitals, inconsistent coding of medical information, and limited awareness about data structure from participating hospital. The pilot phase was completed in 2019 containing

The European medical information framework: A novel ecosystem for sharing healthcare data across Europe

Learn Health Syst. 2019 Dec 25;4(2):e10214. doi: 10.1002/lrh2.10214. eCollection 2020 Apr.

Authors

Simon Lovestone, EMIF Consortium

Abstract

Introduction: The European medical information framework (EMIF) was an Innovative Medicines Initiative project jointly supported by the European Union and the European Federation of Pharmaceutical Industries and Associations, that generated a common technology and governance framework to identify, assess and (re)use healthcare data, to facilitate real-world data research. The objectives of EMIF included providing a unified platform to support a wide range of studies within two verification programmes-Alzheimer's disease (EMIF-AD), and metabolic consequences of obesity (EMIF-MET).

Methods: The EMIF platform was built around two main data-types: electronic health record data and research cohort data, and the platform architecture composed of a set of tools designed to enable data discovery and characterisation. This included the EMIF catalogue, which allowed users

Common data model:
Feeding data into the model

Converting clinical document architecture documents to the common data model for incorporating health information exchange data in observational health studies: CDA to CDM

J Biomed Inform. 2020 Jul;107:103459. doi: 10.1016/j.jbi.2020.103459. Epub 2020 May 26.

Authors

Hyerim Ji, Seok Kim, Soyoung Yi, Hee Hwang, Jeong-Whun Kim, Sooyoung Yoo

Abstract

Background: Utilization of standard health information exchange (HIE) data is growing due to the high adoption rate and interoperability of electronic health record (EHR) systems. However, integration of HIE data into an EHR system is not yet fully adopted in clinical research. In addition, data quality should be verified for the secondary use of these data. Thus, the aims of this study were to convert referral documents in a Health Level 7 (HL7) clinical document architecture (CDA) to the common data model (CDM) to facilitate HIE data availability for longitudinal data analysis, and to identify data quality levels for application in future clinical studies.

Methods: A total of 21,492 referral CDA documents accumulated for over 10 years in a tertiary general hospital in South Korea were analyzed. To convert CDA documents to the Observational

COVID-19 SignSym: A fast adaptation of general clinical NLP tools to identify and normalize COVID-19 signs and symptoms to OMOP common data model

ArXiv. 2020 Jul 13;arXiv:2007.10286v3.Preprint

Authors

Jingqi Wang, Huy Anh Pham, Frank Manion, Masoud Rouhizadeh, Yaoyun Zhang

Abstract

The COVID-19 pandemic swept across the world rapidly infecting millions of people. An efficient tool that can accurately recognize important clinical concepts of COVID-19 from free text in electronic health records will be significantly valuable to accelerate various applications of COVID-19 research. To this end, the existing clinical NLP tool CLAMP was quickly adapted to COVID-19 information and generated an automated tool called COVID-19 SignSym, which can extract and signs/symptoms and their eight attributes such as temporal information and negations from clinical text. The extracted information is also mapped to standard clinical concepts in the common data model of OHDSI OMOP. Evaluation on clinical notes and medical dialogues demonstrated promising results. It is freely accessible to the community as a downloadable package

Common data model: Adding domains

Optimization of Electronic Medical Records for Data Mining Using a Common Data Model

Top Companion Anim Med. 2019 Dec;37:100364. doi: 10.1016/j.tcam.2019.100364. Epub 2019 Sep 26.

Authors

Manlik Kwong, Heather L Gardner, Neil Dieterle, Virginia Rentko

Abstract

The increasing use of electronic health records (EHRs) in veterinary medicine creates an opportunity to utilize the high volume of electronic patient data for mining and data-driven analytics with the goal of improving patient care and outcomes. A central focus of the Clinical and Translational Science Award One Health Alliance (COHA) is to integrate efforts across multiple disciplines to better understand shared diseases in animals and people. The ability to combine veterinary and human medical data provides a unique resource to study the interactions and relationships between animals, humans, and the environment. However, to effectively answer these questions, veterinary EHR data must first be prepared in the same way it is now commonly being done in human medicine to enable data mining and development of analytics to facilitate knowledge formation and solutions that advance our understanding of disease processes, with the

Mapping Local Biospecimen Records to the OMOP Common Data Model

AMIA Jt Summits Transl Sci Proc. 2020 May 30;2020:422-429. eCollection 2020.

Authors

Chelsea L Michael, Evan T Sholle, Regina T Wulff, Gail J Roboz, Thomas R Campion Jr

Abstract

Research to support precision medicine for leukemia patients requires integration of biospecimen and clinical data. The Observational Medical Outcomes Partnership common data model (OMOP CDM) and its Specimen table presents a potential solution. Although researchers have described progress and challenges in mapping electronic health record (EHR) data to populate the OMOP CDM, to our knowledge no studies have described populating the OMOP CDM with biospecimen data. Using biobank data from our institution, we mapped 26% of biospecimen records to the OMOP Specimen table. Records failed mapping due to local codes for time point that were incompatible with the OMOP reference terminology. We recommend expanding allowable codes to encompass research data, adding foreign keys to leverage additional OMOP tables with data from other sources or to store additional specimen details, and considering a new table to represent processed samples and inventory.

Common data model: Supporting pharmacovigilance

Nintedanib and ischemic colitis: Signal assessment with the integrated use of two types of real-world evidence, spontaneous reports of suspected adverse drug reactions, and observational data from large health-care databases

Pharmacoepidemiol Drug Saf. 2020 Aug;29(8):951-957. doi: 10.1002/pds.5022. Epub 2020 May 12.

Author

Rebecca E Chandler

Abstract

Purpose: Statistical screening of Vigibase, the global database of individual case safety reports, highlighted an association between the MedDRA Preferred Term (PT) "colitis" and nintedanib. Nintedanib is a protein kinase inhibitor authorized in accelerated regulatory procedures for the treatment of idiopathic pulmonary fibrosis (IPF). The aim of this report is to describe the integration of two types of real-world evidence, spontaneous reports of adverse drug reactions (ADR), and observational health data (OHD) in the assessment of a post-authorization safety signal

Detecting and Filtering Immune-Related Adverse Events Signal Based on Text Mining and Observational Health Data Sciences and Informatics Common Data Model: Framework Development Study

JMIR Med Inform. 2020 Jun 12;8(6):e17353. doi: 10.2196/17353.

Authors

Yue Yu, Kathryn Ruddy, Aaron Mansfield, Nansu Zong, Andrew Wen, Shintaro Tsuji, Ming Huang, Hongfang Liu, Nilay Shah, Guoqian Jiang

Abstract

Background: Immune checkpoint inhibitors are associated with unique immune-related adverse events (irAEs). As most of the immune checkpoint inhibitors are new to the market, it is important to conduct studies using real-world data sources to investigate their safety profiles.

Objective: The aim of the study was to develop a framework for signal detection and filtration of novel irAEs for 6 Food and Drug Administration-approved immune checkpoint inhibitors.

Empirical assessment of case-based methods for drug safety alert identification in the French National Healthcare System database (SNDS): Methodology of the ALCAPONE project

Pharmacoepidemiol Drug Saf. 2020 Sep;29(9):993-1000. doi: 10.1002/pds.4983. Epub 2020 Mar 4.

Authors

Nicolas H Thurin, Régis Lassalle, Martijn Schuemie, Marine Pénichon, Joshua J Gagne, Jeremy A Rassen, Jacques Benichou, Alain Weill, Patrick Blin, Nicholas Moore, Cécile Droz-Perroteau

Abstract

Objectives: To introduce the methodology of the ALCAPONE project.

Background: The French National Healthcare System Database (SNDS), covering 99% of the French population, provides a potentially valuable opportunity for drug safety alert generation. ALCAPONE aimed to assess empirically in the SNDS case-based designs for alert generation related to four health outcomes of interest.

Methods: ALCAPONE used a reference set adapted from observational medical outcomes

Common data model: Supporting trial recruitment

Automatic trial eligibility surveillance based on unstructured clinical data

Int J Med Inform. 2019 Sep;129:13-19. doi: 10.1016/j.ijmedinf.2019.05.018. Epub 2019 May 23.

Authors

Stéphane M Meystre, Paul M Heider, Youngjun Kim, Daniel B Aruch, Carolyn D Britten

Abstract

Introduction: Insufficient patient enrollment in clinical trials remains a serious and costly problem and is often considered the most critical issue to solve for the clinical trials community. In this project, we assessed the feasibility of automatically detecting a patient's eligibility for a sample of breast cancer clinical trials by mapping coded clinical trial eligibility criteria to the corresponding clinical information automatically extracted from text in the EHR.

Methods: Three open breast cancer clinical trials were selected by oncologists. Their eligibility criteria were manually abstracted from trial descriptions using the OHDSI ATLAS web application. Patients enrolled or screened for these trials were selected as 'positive' or 'possible' cases. Other patients diagnosed with breast cancer were selected as 'negative' cases. A selection of the clinical

Design for a Modular Clinical Trial Recruitment Support System Based on FHIR and OMOP

Stud Health Technol Inform. 2020 Jun 16;270:158-162. doi: 10.3233/SHTI200142.

Authors

Ines Reinecke, Christian Gulden, Michéle Kümmel, Azadeh Nassirian, Romina Blasini, Martin Sedlmayr

Abstract

The MIRACUM consortium is developing a Clinical Trials Recruitment Support System to support the data-driven recruitment of patients for clinical trials. The design of the prototype includes both open source solutions (OMOP CDM, Atlas) and open standards for interoperability (FHIR). The aim of the prototype is to create a patient screening list of potential participants for a clinical study. The paper shows the modular structure and functionality of the prototype building the foundation for the practical implementation of the CTRSS and, at the same time, demonstrating the use of open source solutions and standards for the development of clinical support systems.

International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium

NPJ Digit Med. 2020 Aug 19;3:109. doi: 10.1038/s41746-020-00308-0. eCollection 2020.

Authors

Gabriel A Brat, Griffin M Weber, Nils Gehlenborg, Paul Avillach, Nathan P Palmer, Luca Chiovato, James Cimino, Lemuel R Waitman, Gilbert S Omenn, Alberto Malovini, Jason H Moore, Brett K Beaulieu-Jones, Valentina Tibollo, Shawn N Murphy, Sehi L' Yi, Mark S Keller, Riccardo Bellazzi, David A Hanauer, Arnaud Serret-Larmande, Alba Gutierrez-Sacristan, John J Holmes, Douglas S Bell, Kenneth D Mandl, Robert W Follett, Jeffrey G Klann, Douglas A Murad, Luigia Scudeller, Mauro Bucalo, Katie Kirchoff, Jean Craig, Jihad Obeid, Vianney Jouhet, Romain Griffier, Sebastien Cossin, Bertrand Moal, Lav P Patel, Antonio Bellasi, Hans U Prokosch, Detlef Kraska, Piotr Sliz, Amelia L M Tan, Kee Yuan Ngiam, Alberto Zambelli, Danielle L Mowery, Emily Schiver, Batsal Devkota, Robert L Bradford, Mohamad Daniar, Christel Daniel, Vincent Benoit, Romain Bey, Nicolas Paris, Patricia Serre, Nina Orlova, Julien Dubiel, Martin Hilka, Anne Sophie Jannot, Stephane Breant, Judith Leblanc, Nicolas Griffon, Anita Burgun, Melodie Bernaux, Arnaud Sandrin, Elisa Salamanca, Sylvie Cormont, Thomas Ganslandt, Tobias Gradinger, Julien Champ, Martin Boeker, Patricia Martel, Loic Esteve, Alexandre Gramfort, Olivier Grisel, Damien Leprovost, Thomas Moreau, Gael Varoquaux, Jill-Jênn Vie, Demian Wassermann, Arthur Mensch,

Translating evidence into practice: eligibility criteria fail to eliminate clinically significant differences between real-world and study populations

NPJ Digit Med. 2020 May 11;3:67. doi: 10.1038/s41746-020-0277-8. eCollection 2020.

Authors

Amelia J Averitt, Chunhua Weng, Patrick Ryan, Adler Perotte

Abstract

Randomized controlled trials (RCTs) are regarded as the most reputable source of evidence. In some studies, factors beyond the intervention itself may contribute to the measured effect, an occurrence known as heterogeneity of treatment effect (HTE). If the RCT population differs from the real-world population on factors that induce HTE, the trials effect will not replicate. The RCTs eligibility criteria should identify the sub-population in which its evidence will replicate. However, the extent to which the eligibility criteria identify the appropriate population is unknown, which raises concerns for generalizability. We compared reported data from RCTs with real-world data from the electronic health records of a large, academic medical center that was curated according to RCT eligibility criteria. Our results show fundamental differences between the RCT population and

Common data model: Surveys

Clinical Research Informatics

Yearb Med Inform. 2020 Aug;29(1):203-207. doi: 10.1055/s-0040-1702007. Epub 2020 Aug 21.

Authors

Christel Daniel, Dipak Kalra, Section Editors for the IMIA Yearbook Section on Clinical Research Informatics

Abstract

Objectives: To summarize key contributions to current research in the field of Clinical Research Informatics (CRI) and to select best papers published in 2019.

Method: A bibliographic search using a combination of MeSH descriptors and free-text terms on CRI was performed using PubMed, followed by a double-blind review in order to select a list of candidate best papers to be then peer-reviewed by external reviewers. After peer-review ranking, a consensus meeting between the two section editors and the editorial team was organized to finally conclude on the selected three best papers.

Results: Among the 517 papers, published in 2019, returned by the search, that were in the scope of the various areas of CRI, the full review process selected three best papers. The first best paper describes the use of a homomorphic encryption technique to enable federated analysis of real-world

Common data model:
Data quality

Extending an open-source tool to measure data quality: case report on Observational Health Data Science and Informatics (OHDSI)

BMJ Health Care Inform. 2020 Mar;27(1):e100054. doi: 10.1136/bmjhci-2019-100054.

Authors

Brian E Dixon, Chen Wen, Tony French, Jennifer L Williams, Jon D Duke, Shaun J Grannis

Abstract

Introduction: As the health system seeks to leverage large-scale data to inform population outcomes, the informatics community is developing tools for analysing these data. To support data quality assessment within such a tool, we extended the open-source software Observational Health Data Sciences and Informatics (OHDSI) to incorporate new functions useful for population health.

Methods: We developed and tested methods to measure the completeness, timeliness and entropy of information. The new data quality methods were applied to over 100 million clinical messages received from emergency department information systems for use in public health syndromic surveillance systems.

Incrementally Transforming Electronic Medical Records into the Observational Medical Outcomes Partnership Common Data Model: A Multidimensional Quality Assurance Approach

Appl Clin Inform. 2019 Oct;10(5):794-803. doi: 10.1055/s-0039-1697598. Epub 2019 Oct 23.

Authors

Kristine E Lynch, Stephen A Deppen, Scott L DuVall, Benjamin Viernes, Aize Cao, Daniel Park, Elizabeth Hanchrow, Kushan Hewa, Peter Greaves, Michael E Matheny

Abstract

Background: The development and adoption of health care common data models (CDMs) has addressed some of the logistical challenges of performing research on data generated from disparate health care systems by standardizing data representations and leveraging standardized terminology to express clinical information consistently. However, transforming a data system into a CDM is not a trivial task, and maintaining an operational, enterprise capable CDM that is incrementally updated within a data warehouse is challenging.

Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model?: A Validation Study Based on Prescribing Codeine in Children

Clin Pharmacol Ther. 2020 Apr;107(4):915-925. doi: 10.1002/cpt.1785. Epub 2020 Mar 2.

Authors

Gianmario Candore, Karin Hedenmalm, Jim Slattery, Alison Cave, Xavier Kurz, Peter Arlett

Abstract

Exploring and combining results from more than one real-world data (RWD) source might be necessary in order to explore variability and demonstrate generalizability of the results or for regulatory requirements. However, the heterogeneous nature of RWD poses challenges when working with more than one source, some of which can be solved by analyzing databases converted into a common data model (CDM). The main objective of the study was to evaluate the implementation of the Observational Medical Outcome Partnership (OMOP) CDM on IQVIA Medical Research Data (IMRD)-UK data. A drug utilization study describing the prescribing of codeine for pain in children was used as a case study to be replicated in IMRD-UK and its

Phenotyping: Methods

Development and validation of phenotype classifiers across multiple sites in the observational health data sciences and informatics network

J Am Med Inform Assoc. 2020 Jun 1;27(6):877-883. doi: 10.1093/jamia/ocaa032.

Authors

Mehr Kashyap, Martin Seneviratne, Juan M Banda, Thomas Falconer, Borim Ryu, Sooyoung Yoo, George Hripcsak, Nigam H Shah

Abstract

Objective: Accurate electronic phenotyping is essential to support collaborative observational research. Supervised machine learning methods can be used to train phenotype classifiers in a high-throughput manner using imperfectly labeled data. We developed 10 phenotype classifiers using this approach and evaluated performance across multiple sites within the Observational Health Data Sciences and Informatics (OHDSI) network.

Materials and methods: We constructed classifiers using the Automated PHenotype Routine for Observational Definition, Identification, Training and Evaluation (APHRODITE) R-package, an open-source framework for learning phenotype classifiers using datasets in the Observational

Adapting electronic health records-derived phenotypes to claims data: Lessons learned in using limited clinical data for phenotyping

J Biomed Inform. 2020 Feb;102:103363. doi: 10.1016/j.jbi.2019.103363. Epub 2019 Dec 19.

Authors

Anna Ostropolets, Christian Reich, Patrick Ryan, Ning Shang, George Hripcsak, Chunhua Weng

Abstract

Algorithms for identifying patients of interest from observational data must address missing and inaccurate data and are desired to achieve comparable performance on both administrative claims and electronic health records data. However, administrative claims data do not contain the necessary information to develop accurate algorithms for disorders that require laboratory results, and this omission can result in insensitive diagnostic code-based algorithms. In this paper, we tested our assertion that the performance of a diagnosis code-based algorithm for chronic kidney disorder (CKD) can be improved by adding other codes indirectly related to CKD (e.g., codes for dialysis, kidney transplant, suspicious kidney disorders). Following the best practices from Observational Health Data Sciences and Informatics (OHDSI), we adapted an electronic health record-based gold standard algorithm for CKD and then created algorithms that can be executed on

PheValuator: Development and evaluation of a phenotype algorithm evaluator

J Biomed Inform. 2019 Sep;97:103258. doi: 10.1016/j.jbi.2019.103258. Epub 2019 Jul 29.

Authors

Joel N Swerdel, George Hripcsak, Patrick B Ryan

Abstract

Background: The primary approach for defining disease in observational healthcare databases is to construct phenotype algorithms (PAs), rule-based heuristics predicated on the presence, absence, and temporal logic of clinical observations. However, a complete evaluation of PAs, i.e., determining sensitivity, specificity, and positive predictive value (PPV), is rarely performed. In this study, we propose a tool (PheValuator) to efficiently estimate a complete PA evaluation.

Methods: We used 4 administrative claims datasets: OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN); IBM MarketScan Multi-State Medicaid); IBM MarketScan Medicare Supplemental Beneficiaries; and IBM MarketScan Commercial Claims and Encounters from 2000 to 2017. Using PheValuator involves (1) creating a diagnostic predictive model for the phenotype, (2) applying the model to a large set of randomly selected subjects, and (3) comparing each

Research and Reporting Considerations for Observational Studies Using Electronic Health Record Data

Ann Intern Med. 2020 Jun 2;172(11_Supplement):S79-S84. doi: 10.7326/M19-0873.

Authors

Alison Callahan, Nigam H Shah, Jonathan H Chen

Abstract

Electronic health records (EHRs) are an increasingly important source of real-world health care data for observational research. Analyses of data collected for purposes other than research require careful consideration of data quality as well as the general research and reporting principles relevant to observational studies. The core principles for observational research in general also apply to observational research using EHR data, and these are well addressed in prior literature and guidelines. This article provides additional recommendations for EHR-based research. Considerations unique to EHR-based studies include assessment of the accuracy of computer-executable cohort definitions that can incorporate unstructured data from clinical notes and management of data challenges, such as irregular sampling, missingness, and variation across time and place. Principled application of existing research and reporting guidelines alongside these additional considerations will improve the quality of EHR-based observational studies.

Deep phenotyping: Embracing complexity and temporality-Towards scalability, portability, and interoperability

J Biomed Inform. 2020 May;105:103433. doi: 10.1016/j.jbi.2020.103433. Epub 2020 Apr 23.

Authors

Chunhua Weng, Nigam H Shah, George Hripcsak

Phenotyping:
Specific phenotypes

Toolkit to Compute Time-Based Elixhauser Comorbidity Indices and Extension to Common Data Models

Healthc Inform Res. 2020 Jul;26(3):193-200. doi: 10.4258/hir.2020.26.3.193. Epub 2020 Jul 31.

Authors

Shorabuddin Syed, Ahmad Baghal, Fred Prior, Meredith Zozus, Shaymaa Al-Shukri, Hafsa Bareen Syeda, Maryam Garza, Salma Begum, Kim Gates, Mahanazuddin Syed, Kevin W Sexton

Abstract

Objective: The time-dependent study of comorbidities provides insight into disease progression and trajectory. We hypothesize that understanding longitudinal disease characteristics can lead to more timely intervention and improve clinical outcomes. As a first step, we developed an efficient and easy-to-install toolkit, the Time-based Elixhauser Comorbidity Index (TECI), which pre-calculates time-based Elixhauser comorbidities and can be extended to common data models (CDMs).

Methods: A Structured Query Language (SQL)-based toolkit, TECI, was built to pre-calculate time-specific Elixhauser comorbidity indices using data from a clinical data repository (CDR). Then it was extended to the Informatics for Integrating Biology and the Bedside (I2B2) and

Supplementing claims data analysis using self-reported data to develop a probabilistic phenotype model for current smoking status

J Biomed Inform. 2019 Sep;97:103264. doi: 10.1016/j.jbi.2019.103264. Epub 2019 Aug 3.

Authors

Jenna M Reps, Peter R Rijnbeek, Patrick B Ryan

Abstract

Objectives: Smoking status is poorly record in US claims data. IBM MarketScan Commercial is a claims database that can be linked to an additional health risk assessment with self-reported smoking status for a subset of 1,966,174 patients. We investigate whether this subset could be used to learn a smoking status phenotype model generalizable to all US claims data that calculates the probability of being a current smoker.

Methods: 251,643 (12.8%) had self-reported their smoking status as 'current smoker'. A regularized logistic regression model, the Current Risk of Smoking Status (CROSS), was trained using the subset of patients with self-reported smoking status. CROSS considered 53,027 candidate covariates including demographics and conditions/drugs/measurements/procedures/observations

Developing Predictive Models to Determine Patients in End-of-Life Care in Administrative Datasets

Drug Saf. 2020 May;43(5):447-455. doi: 10.1007/s40264-020-00906-7.

Authors

Joel N Swerdel, Jenna M Reps, Daniel Fife, Patrick B Ryan

Abstract

Introduction: In observational studies with mortality endpoints, one needs to consider how to account for subjects whose interventions appear to be part of 'end-of-life' care.

Objective: The objective of this study was to develop a diagnostic predictive model to identify those in end-of-life care at the time of a drug exposure.

Methods: We used data from four administrative claims datasets from 2000 to 2017. The index date was the date of the first prescription for the last new drug subjects received during their observation period. The outcome of end-of-life care was determined by the presence of one or more codes indicating terminal or hospice care. Models were developed using regularized logistic regression. Internal validation was through examination of the area under the receiver operating

Identifying the DEAD: Development and Validation of a Patient-Level Model to Predict Death Status in Population-Level Claims Data

Drug Saf. 2019 Nov;42(11):1377-1386. doi: 10.1007/s40264-019-00827-0.

Authors

Jenna M Reps, Peter R Rijnbeek, Patrick B Ryan

Abstract

Introduction: US claims data contain medical data on large heterogeneous populations and are excellent sources for medical research. Some claims data do not contain complete death records, limiting their use for mortality or mortality-related studies. A model to predict whether a patient died at the end of the follow-up time (referred to as the end of observation) is needed to enable mortality-related studies.

Objective: The objective of this study was to develop a patient-level model to predict whether the end of observation was due to death in US claims data.

Methods: We used a claims dataset with full death records, Optum De-Identified Clinformatics

Phenotyping: Process

Making work visible for electronic phenotype implementation: Lessons learned from the eMERGE network

J Biomed Inform. 2019 Nov;99:103293. doi: 10.1016/j.jbi.2019.103293. Epub 2019 Sep 19.

Authors

Ning Shang, Cong Liu, Luke V Rasmussen, Casey N Ta, Robert J Caroll, Barbara Benoit, Todd Lingren, Ozan Dikilitas, Frank D Mentch, David S Carrell, Wei-Qi Wei, Yuan Luo, Vivian S Gainer, Iftikhar J Kullo, Jennifer A Pacheco, Hakon Hakonarson, Theresa L Walunas, Joshua C Denny, Ken Wiley, Shawn N Murphy, George Hripcsak, Chunhua Weng

Abstract

Background: Implementation of phenotype algorithms requires phenotype engineers to interpret human-readable algorithms and translate the description (text and flowcharts) into computable phenotypes - a process that can be labor intensive and error prone. To address the critical need for reducing the implementation efforts, it is important to develop portable algorithms.

Methods: We conducted a retrospective analysis of phenotype algorithms developed in the Electronic Medical Records and Genomics (eMERGE) network and identified common customization tasks required for implementation. A novel scoring system was developed to

Considerations for Improving the Portability of Electronic Health Record-Based Phenotype Algorithms

AMIA Annu Symp Proc. 2020 Mar 4;2019:755-764. eCollection 2019.

Authors

Luke V Rasmussen, Pascal S Brandt, Guoqian Jiang, Richard C Kiefer, Jennifer A Pacheco, Prakash Adekkanattu, Jessica S Ancker, Fei Wang, Zhenxing Xu, Jyotishman Pathak, Yuan Luo

Abstract

With the increased adoption of electronic health records, data collected for routine clinical care is used for health outcomes and population sciences research, including the identification of phenotypes. In recent years, research networks, such as eMERGE, OHDSI and PCORnet, have been able to increase statistical power and population diversity by combining patient cohorts. These networks share phenotype algorithms that are executed at each participating site. Here we observe experiences with phenotype algorithm portability across seven research networks and propose a generalizable framework for phenotype algorithm portability. Several strategies exist to increase the portability of phenotype algorithms, reducing the implementation effort needed by each site. These include using a common data model, standardized representation of the phenotype algorithm logic, and technical solutions to facilitate federated execution of queries. Portability is achieved by

Phenotyping:
Tools for using the CDM

PatientExploreR: an extensible application for dynamic visualization of patient clinical history from electronic health records in the OMOP common data model

Bioinformatics. 2019 Nov 1;35(21):4515-4518. doi: 10.1093/bioinformatics/btz409.

Authors

Benjamin S Glicksberg, Boris Oskotsky, Phyllis M Thangaraj, Nicholas Giangreco, Marcus A Badgeley, Kipp W Johnson, Debajyoti Datta, Vivek A Rudrapatna, Nadav Rappoport, Mark M Shervey, Riccardo Miotto, Theodore C Goldstein, Eugenia Rutenberg, Remi Frazier, Nelson Lee, Sharat Israni, Rick Larsen, Bethany Percha, Li Li, Joel T Dudley, Nicholas P Tatonetti, Atul J Butte

Abstract

Motivation: Electronic health records (EHRs) are quickly becoming omnipresent in healthcare, but interoperability issues and technical demands limit their use for biomedical and clinical research. Interactive and flexible software that interfaces directly with EHR data structured around a common data model (CDM) could accelerate more EHR-based research by making the data more accessible to researchers who lack computational expertise and/or domain knowledge.

KETOS: Clinical decision support and machine learning as a service - A training and deployment platform based on Docker, OMOP-CDM, and FHIR Web Services

PLoS One. 2019 Oct 3;14(10):e0223010. doi: 10.1371/journal.pone.0223010. eCollection 2019.

Authors

Julian Gruendner, Thorsten Schwachhofer, Phillip Sippl, Nicolas Wolf, Marcel Erpenbeck, Christian Gulden, Lorenz A Kapsner, Jakob Zierk, Sebastian Mate, Michael Stürzl, Roland Croner, Hans-Ulrich Prokosch, Dennis Toddenroth

Abstract

Background and objective: To take full advantage of decision support, machine learning, and patient-level prediction models, it is important that models are not only created, but also deployed in a clinical setting. The KETOS platform demonstrated in this work implements a tool for researchers allowing them to perform statistical analyses and deploy resulting models in a secure environment.

Methods: The proposed system uses Docker virtualization to provide researchers with reproducible

EHR-Independent Predictive Decision Support Architecture Based on OMOP

Appl Clin Inform. 2020 May;11(3):399-404. doi: 10.1055/s-0040-1710393. Epub 2020 Jun 3.

Authors

Philipp Unberath, Hans Ulrich Prokosch, Julian Gründner, Marcel Erpenbeck, Christian Maier, Jan Christoph

Abstract

Background: The increasing availability of molecular and clinical data of cancer patients combined with novel machine learning techniques has the potential to enhance clinical decision support, example, for assessing a patient's relapse risk. While these prediction models often produce promising results, a deployment in clinical settings is rarely pursued.

Objectives: In this study, we demonstrate how prediction tools can be integrated generically into a clinical setting and provide an exemplary use case for predicting relapse risk in melanoma patients.

Methods: To make the decision support architecture independent of the electronic health record (EHR) and transferable to different hospital environments, it was based on the widely used

Characterization

Incidence, prevalence and prescription patterns of antipsychotic medications use in Asia and US: A cross-nation comparison with common data model

J Psychiatr Res. 2020 Aug 27;131:77-84. doi: 10.1016/j.jpsychires.2020.08.025. Online ahead of print.

Authors

Chien-Chou Su, Edward Chia-Cheng Lai, Yea-Huei Kao Yang, Kenneth K C Man, Kiyoshi Kubota, Paul Stang, Martijn Schuemie, Patrick Ryan, Chantelle Hardy, Yinghong Zhang, Shinya Kimura, Yukari Kamijima, IanC K Wong, Soko Setoguchi

Abstract

The use of antipsychotic medications (APMs) could be different among countries due to availability, approved indications, characteristics and clinical practice. However, there is limited literature providing comparisons of APMs use among countries. To examine trends in antipsychotic prescribing in Taiwan, Hong Kong, Japan, and the United States, we conducted a cross-national study from 2002 to 2014 b y using the distributed network approach with common data model. We included all patients who had at least a record of antipsychotic prescription in this study, and defined patients without previous exposure of antipsychotics for 6 months before the index date as

An international characterisation of patients hospitalised with COVID-19 and a comparison with those previously hospitalised with influenza

Nature Communications and medRxiv. 2020 Apr 25;2020.04.22.20074336. doi: 10.1101/2020.04.22.20074336. Preprint

Authors

Edward Burn, Seng Chan You, Anthony G Sena, Kristin Kostka, Hamed Abedtash, Maria Tereza F Abrahão, Amanda Alberga, Heba Alghoul, Osaid Alser, Thamir M Alshammari, Carlos Areia, Juan M Banda, Jaehyeong Cho, Aedin C Culhane, Alexander Davydov, Frank J DeFalco, Talita Duarte-Salles, Scott DuVall, Thomas Falconer, Weihua Gao, Asieh Golozar, Jill Hardin, George Hripcsak, Vojtech Huser, Hokyun Jeon, Yonghua Jing, Chi Young Jung, Benjamin Skov Kaas-Hansen, Denys Kaduk, Seamus Kent, Yeesuk Kim, Spyros Kolovos, Jennifer C E Lane, Hyejin Lee, Kristine E Lynch, Rupa Makadia, Michael E Matheny, Paras Mehta, Daniel R Morales, Karthik Natarajan, Fredrik Nyberg, Anna Ostropolets, Rae Woong Park, Jimyung Park, Jose D Posada, Albert Prats-Uribe, Gowtham Rao, Christian Reich, Yeunsook Rho, Peter Rijnbeek, Selva Muthu Kumaran Sathappan, Lisa M Schilling, Martijn Schuemie, Nigam H Shah, Azza Shoaibi, Seokyoung Song, Matthew Spotnitz, Marc A Suchard, Joel N Swerdel, David Vizcaya, Salvatore Volpe, Haini Wen, Andrew E Williams, Belay B Yimer, Lin Zhang, Oleg Zhuk, Daniel Prieto-Alhambra, Patrick Ryan

Comparison of the incidence of diabetes in United States and Indian youth: An international harmonization of youth diabetes registries

Pediatr Diabetes. 2020 Mar 20. doi: 10.1111/pedi.13009. Online ahead of print.

Authors

Elizabeth T Jensen, Dana A Dabelea, Pradeep A Praveen, Anandakumar Amutha, Christine W Hockett, Scott P Isom, Toan C Ong, Viswanathan Mohan, Ralph D'Agostino Jr, Michael G Kahn, Richard F Hamman, Paul Wadwa, Lawrence Dolan, Jean M Lawrence, S V Madhu, Reshmi Chhokar, Komal Goel, Nikhil Tandon, Elizabeth Mayer-Davis

Abstract

Objective: Incidence of youth-onset diabetes in India has not been well described. Comparison of incidence, across diabetes registries, has the potential to inform hypotheses for risk factors. We sought to compare the incidence of diabetes in the U.S.-based registry of youth onset diabetes (SEARCH) to the Registry of Diabetes with Young Age at Onset (YDR-Chennai and New Delhi regions) in India.

Methods: We harmonized data from both SEARCH and YDR to the Observational Medical

Characterization of Anti-seizure Medication Treatment Pathways in Pediatric Epilepsy Using the Electronic Health Record-Based Common Data Model

Front Neurol. 2020 May 12;11:409. doi: 10.3389/fneur.2020.00409. eCollection 2020.

Authors

Hunmin Kim, Sooyoung Yoo, Yonghoon Jeon, Soyoung Yi, Seok Kim, Sun Ah Choi, Hee Hwang, Ki Joong Kim

Abstract

The purpose of this pilot study was to analyze treatment pathways of pediatric epilepsy using the common data model (CDM) based on electronic health record (EHR) data. We also aimed to reveal whether CDM analysis was feasible and applicable to epilepsy research. We analyzed the treatment pathways of pediatric epilepsy patients from our institute who underwent antiseizure medication (ASM) treatment for at least 2 years, using the Observational Medical Outcomes Partnership (OMOP)-CDM. Subgroup analysis was performed for generalized or focal epilepsy, varying age of epilepsy onset, and specific epilepsy syndromes. Changes in annual prescription patterns were also analyzed to reveal the different trends. We also calculated the proportion of drug-resistant epilepsy

Imputation and characterization of uncoded self-harm in major mental illness using machine learning

J Am Med Inform Assoc. 2020 Jan 1;27(1):136-146. doi: 10.1093/jamia/ocz173.

Authors

Praveen Kumar, Anastasiya Nestsiarovich, Stuart J Nelson, Berit Kerner, Douglas J Perkins, Christophe G Lambert

Abstract

Objective: We aimed to impute uncoded self-harm in administrative claims data of individuals with major mental illness (MMI), characterize self-harm incidence, and identify factors associated with coding bias.

Materials and methods: The IBM MarketScan database (2003-2016) was used to analyze visit-level self-harm in 10 120 030 patients with ≥2 MMI codes. Five machine learning (ML) classifiers were tested on a balanced data subset, with XGBoost selected for the full dataset. Classification performance was validated via random data mislabeling and comparison with a clinician-derived "gold standard." The incidence of coded and imputed self-harm was characterized by year, patient age, sex, U.S. state, and MMI diagnosis.

Application of a Common Data Model (CDM) to rank the paediatric user and prescription prevalence of 15 different drug classes in South Korea, Hong Kong, Taiwan, Japan and Australia: an observational, descriptive study

BMJ Open. 2020 Jan 13;10(1):e032426. doi: 10.1136/bmjopen-2019-032426.

Authors

Ruth Brauer, Ian Chi Kei Wong, Kenneth Kc Man, Nicole L Pratt, Rae Woong Park, Soo-Yeon Cho, Yu-Chuan Jack Li, Usman Iqbal, Phung-Anh Alex Nguyen, Martijn Schuemie

Abstract

Objective: To measure the paediatric user and prescription prevalence in inpatient and ambulatory settings in South Korea, Hong Kong, Taiwan, Japan and Australia by age and gender. A further objective was to list the most commonly used drugs per drug class, per country.

Design and setting: Hospital inpatient and insurance paediatric healthcare data from the following databases were used to conduct this descriptive drug utilisation study: (i) the South Korean Ajou University School of Medicine database; (ii) the Hong Kong Clinical Data Analysis and Reporting

Treatment Patterns for Chronic Comorbid Conditions in Patients With Cancer Using a Large-Scale Observational Data Network

JCO Clin Cancer Inform. 2020 Mar;4:171-183. doi: 10.1200/CCI.19.00107.

Authors

Ruijun Chen, Patrick Ryan, Karthik Natarajan, Thomas Falconer, Katherine D Crew, Christian G Reich, Rohit Vashisht, Gurvaneet Randhawa, Nigam H Shah, George Hripcsak

Abstract

Purpose: Patients with cancer are predisposed to developing chronic, comorbid conditions that affect prognosis, quality of life, and mortality. While treatment guidelines and care variations for these comorbidities have been described for the general noncancer population, less is known about real-world treatment patterns in patients with cancer. We sought to characterize the prevalence and distribution of initial treatment patterns across a large-scale data network for depression, hypertension, and type II diabetes mellitus (T2DM) among patients with cancer.

Methods: We used the Observational Health Data Sciences and Informatics network, an international collaborative implementing the Observational Medical Outcomes Partnership

Estimation: Methods

A plea to stop using the case-control design in retrospective database studies

Stat Med. 2019 Sep 30;38(22):4199-4208. doi: 10.1002/sim.8215. Epub 2019 Aug 22.

Authors

Martijn J Schuemie, Patrick B Ryan, Kenneth K C Man, Ian C K Wong, Marc A Suchard, George Hripcsak

Abstract

The case-control design is widely used in retrospective database studies, often leading to spectacular findings. However, results of these studies often cannot be replicated, and the advantage of this design over others is questionable. To demonstrate the shortcomings of applications of this design, we replicate two published case-control studies. The first investigates isotretinoin and ulcerative colitis using a simple case-control design. The second focuses on dipeptidyl peptidase-4 inhibitors and acute pancreatitis, using a nested case-control design. We include large sets of negative control exposures (where the true odds ratio is believed to be 1) in both studies. Both replication studies produce effect size estimates consistent with the original studies, but also generate estimates for the negative control exposures showing substantial residual bias. In contrast, applying a self-controlled design to answer the same questions using the same

Empirical assessment of case-based methods for identification of drugs associated with upper gastrointestinal bleeding in the French National Healthcare System database (SNDS)

Pharmacoepidemiol Drug Saf. 2020 Aug;29(8):890-903. doi: 10.1002/pds.5038. Epub 2020 Jun 10.

Authors

Nicolas H Thurin, Régis Lassalle, Martijn Schuemie, Marine Pénichon, Joshua J Gagne, Jeremy A Rassen, Jacques Benichou, Alain Weill, Patrick Blin, Nicholas Moore, Cécile Droz-Perroteau

Abstract

Purpose: Upper gastrointestinal bleeding (UGIB) is a severe and frequent drug-related event. In order to enable efficient drug safety alert generation in the French National Healthcare System database (SNDS), we assessed and calibrated empirically case-based designs to identify drug associated with UGIB risk.

Methods: All cases of UGIB were extracted from SNDS (2009-2014) using two definitions. Positive and negative drug controls were used to compare 196 self-controlled case series (SCCS),

Robust-ODAL: Learning from heterogeneous health systems without sharing patient-level data

Pac Symp Biocomput. 2020;25:695-706.

Authors

Jiayi Tong, Rui Duan, Ruowang Li, Martijn J Scheuemie, Jason H Moore, Yong Chen

Abstract

Electronic Health Records (EHR) contain extensive patient data on various health outcomes and risk predictors, providing an efficient and wide-reaching source for health research. Integrated EHR data can provide a larger sample size of the population to improve estimation and prediction accuracy. To overcome the obstacle of sharing patient-level data, distributed algorithms were developed to conduct statistical analyses across multiple clinical sites through sharing only aggregated information. However, the heterogeneity of data across sites is often ignored by existing distributed algorithms, which leads to substantial bias when studying the association between the outcomes and exposures. In this study, we propose a privacy-preserving and communication-efficient distributed algorithm which accounts for the heterogeneity caused by a small number of the clinical sites. We evaluated our algorithm through a systematic simulation study motivated by real-world scenarios and applied our algorithm to multiple claims datasets from the Observational

Principles of Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND)

J Am Med Inform Assoc. 2020 Aug 1;27(8):1331-1337. doi: 10.1093/jamia/ocaa103.

Authors

Martijn J Schuemie, Patrick B Ryan, Nicole Pratt, RuiJun Chen, Seng Chan You, Harlan M Krumholz, David Madigan, George Hripcsak, Marc A Suchard

Abstract

Evidence derived from existing health-care data, such as administrative claims and electronic health records, can fill evidence gaps in medicine. However, many claim such data cannot be used to estimate causal treatment effects because of the potential for observational study bias; for example, due to residual confounding. Other concerns include P hacking and publication bias. In response, the Observational Health Data Sciences and Informatics international collaborative launched the Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) research initiative. Its mission is to generate evidence on the effects of medical interventions using observational health-care databases while addressing the aforementioned concerns by following a recently proposed paradigm. We define 10 principles of LEGEND that enshrine this new paradigm, prescribing the generation and dissemination of evidence on many

Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using hypertension as a case study

J Am Med Inform Assoc. 2020 Aug 1;27(8):1268-1277. doi: 10.1093/jamia/ocaa124.

Authors

Martijn J Schuemie, Patrick B Ryan, Nicole Pratt, RuiJun Chen, Seng Chan You, Harlan M Krumholz, David Madigan, George Hripcsak, Marc A Suchard

Abstract

Objectives: To demonstrate the application of the Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) principles described in our companion article to hypertension treatments and assess internal and external validity of the generated evidence.

Materials and methods: LEGEND defines a process for high-quality observational research based on 10 guiding principles. We demonstrate how this process, here implemented through large-scale propensity score modeling, negative and positive control questions, empirical calibration, and full transparency, can be applied to compare antihypertensive drug therapies. We assess internal validity through covariate balance, confidence-interval coverage, between-database heterogeneity,

Channeling Bias in the Analysis of Risk of Myocardial Infarction, Stroke, Gastrointestinal Bleeding, and Acute Renal Failure with the Use of Paracetamol Compared with Ibuprofen

Drug Saf. 2020 Sep;43(9):927-942. doi: 10.1007/s40264-020-00950-3.

Authors

Rachel B Weinstein, Patrick B Ryan, Jesse A Berlin, Martijn J Schuemie, Joel Swerdel, Daniel Fife

Abstract

Introduction: Observational studies estimating severe outcomes for paracetamol versus ibuprofen use have acknowledged the specific challenge of channeling bias. A previous study relying on negative controls suggested that using large-scale propensity score (LSPS) matching may mitigate bias better than models using limited lists of covariates.

Objective: The aim was to assess whether using LSPS matching would enable the evaluation of paracetamol, compared to ibuprofen, and increased risk of myocardial infarction, stroke, gastrointestinal (GI) bleeding, or acute renal failure.

Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence

JAMA Netw Open. 2019 Oct 2;2(10):e1912869. doi: 10.1001/jamanetworkopen.2019.12869.

Authors

Victoria L Bartlett, Sanket S Dhruva, Nilay D Shah, Patrick Ryan, Joseph S Ross

Abstract

Importance: Although randomized clinical trials are considered to be the criterion standard for generating clinical evidence, the use of real-world evidence to evaluate the efficacy and safety of medical interventions is gaining interest. Whether observational data can be used to address the same clinical questions being answered by traditional clinical trials is still unclear.

Objective: To identify the number of clinical trials published in high-impact journals in 2017 that could be feasibly replicated using observational data from insurance claims and/or electronic health records (EHRs).

Design, setting, and participants: In this cross-sectional analysis, PubMed was searched to identify all US-based clinical trials, regardless of randomization, published between January 1,

The Counterfactual χ-GAN: Finding comparable cohorts in observational health data

J Biomed Inform. 2020 Sep;109:103515. doi: 10.1016/j.jbi.2020.103515. Epub 2020 Aug 7.

Authors

Amelia J Averitt, Natnicha Vanitchanant, Rajesh Ranganath, Adler J Perotte

Abstract

Causal inference often relies on the counterfactual framework, which requires that treatment assignment is independent of the outcome, known as strong ignorability. Approaches to enforcing strong ignorability in causal analyses of observational data include weighting and matching methods. Effect estimates, such as the average treatment effect (ATE), are then estimated as expectations under the re-weighted or matched distribution, P. The choice of P is important and can impact the interpretation of the effect estimate and the variance of effect estimates. In this work, instead of specifying P, we learn a distribution that simultaneously maximizes coverage and minimizes variance of ATE estimates. In order to learn this distribution, this research proposes a generative adversarial network (GAN)-based model called the Counterfactual χ-GAN (cGAN), which also learns feature-balancing weights and supports unbiased causal estimation in the absence

Estimation: Results

Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

Lancet. 2019 Nov 16;394(10211):1816-1826. doi: 10.1016/S0140-6736(19)32317-7. Epub 2019 Oct 24.

Authors

Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, RuiJun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David Madigan, George Hripcsak, Patrick B Ryan

Abstract

Background: Uncertainty remains about the optimal monotherapy for hypertension, with current guidelines recommending any primary agent among the first-line drug classes thiazide or thiazide-like diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, and non-dihydropyridine calcium channel blockers, in the absence of comorbid indications. Randomised trials have not further refined this choice.

Methods: We developed a comprehensive framework for real-world evidence that enables comparative effectiveness and safety evaluation across many drugs and outcomes from

Comparison of Cardiovascular and Safety Outcomes of Chlorthalidone vs Hydrochlorothiazide to Treat Hypertension

JAMA Intern Med. 2020 Apr 1;180(4):542-551. doi: 10.1001/jamainternmed.2019.7454.

Authors

George Hripcsak, Marc A Suchard, Steven Shea, RuiJun Chen, Seng Chan You, Nicole Pratt, David Madigan, Harlan M Krumholz, Patrick B Ryan, Martijn J Schuemie

Abstract

Importance: Chlorthalidone is currently recommended as the preferred thiazide diuretic to treat hypertension, but no trials have directly compared risks and benefits.

Objective: To compare the effectiveness and safety of chlorthalidone and hydrochlorothiazide as first-line therapies for hypertension in real-world practice.

Design, setting, and participants: This is a Large-Scale Evidence Generation and Evaluation in a Network of Databases (LEGEND) observational comparative cohort study with large-scale propensity score stratification and negative-control and synthetic positive-control calibration on

Comparison of First-Line Dual Combination Treatments in Hypertension: Real-World Evidence from Multinational Heterogeneous Cohorts

Korean Circ J. 2020 Jan;50(1):52-68. doi: 10.4070/kcj.2019.0173. Epub 2019 Aug 28.

Authors

Seng Chan You, Sungjae Jung, Joel N Swerdel, Patrick B Ryan, Martijn J Schuemie, Marc A Suchard, Seongwon Lee, Jaehyeong Cho, George Hripcsak, Rae Woong Park, Sungha Park

Abstract

Background and objectives: 2018 ESC/ESH Hypertension guideline recommends 2-drug combination as initial anti-hypertensive therapy. However, real-world evidence for effectiveness of recommended regimens remains limited. We aimed to compare the effectiveness of first-line anti-hypertensive treatment combining 2 out of the following classes: angiotensin-converting enzyme (ACE) inhibitors/angiotensin-receptor blocker (A), calcium channel blocker (C), and thiazide-type diuretics (D).

Methods: Treatment-naïve hypertensive adults without cardiovascular disease (CVD) who initiated

Comparative safety and effectiveness of alendronate versus raloxifene in women with osteoporosis

Sci Rep. 2020 Jul 6;10(1):11115. doi: 10.1038/s41598-020-68037-8.

Authors

Yeesuk Kim, Yuxi Tian, Jianxiao Yang, Vojtech Huser, Peng Jin, Christophe G Lambert, Hojun Park, Seng Chan You, Rae Woong Park, Peter R Rijnbeek, Mui Van Zandt, Christian Reich, Rohit Vashisht, Yonghui Wu, Jon Duke, George Hripcsak, David Madigan, Nigam H Shah, Patrick B Ryan, Martijn J Schuemie, Marc A Suchard

Abstract

Alendronate and raloxifene are among the most popular anti-osteoporosis medications. However, there is a lack of head-to-head comparative effectiveness studies comparing the two treatments. We conducted a retrospective large-scale multicenter study encompassing over 300 million patients across nine databases encoded in the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The primary outcome was the incidence of osteoporotic hip fracture, while secondary outcomes were vertebral fracture, atypical femoral fracture (AFF), osteonecrosis of the jaw (ONJ), and esophageal cancer. We used propensity score trimming and stratification based on an expansive propensity score model with all pre-treatment patient characteritistcs. We

Effect of Age on the Initiation of Biologic Agent Therapy in Patients With Inflammatory Bowel Disease: Korean Common Data Model Cohort Study

JMIR Med Inform. 2020 Apr 15;8(4):e15124. doi: 10.2196/15124.

Authors

Youn I Choi, Yoon Jae Kim, Jun-Won Chung, Kyoung Oh Kim, Hakki Kim, Rae Woong Park, Dong Kyun Park

Abstract

Background: The Observational Health Data Sciences and Informatics (OHDSI) network is an international collaboration established to apply open-source data analytics to a large network of health databases, including the Korean common data model (K-CDM) network.

Objective: The aim of this study is to analyze the effect that age at diagnosis has on the prognosis of inflammatory bowel disease (IBD) in Korea using a CDM network database.

Methods: We retrospectively analyzed the K-CDM network database from 2005 to 2015. We transformed the electronic medical record into the CDM version 5.0 used in OHDSI. A worsened

Opioid use, postoperative complications, and implant survival after unicompartmental versus total knee replacement: a population-based network study

The Lancet Rheumatology. Volume 1, Issue 4, December 2019, Pages e229-e236

Authors

Edward Burn, James Weaver, Daniel Morales, Albert Prats-Uribe, Antonella Delmestri, Victoria Y Strauss, Ying He, Danielle E Robinson, Rafael Pinedo-Villanueva, Spyros Kolovos, Talita Duarte-Salles, William Sproviero, Dahai Yu, Michel Van Speybroeck, Ross Williams, Luis H John, Nigel Hughes, Anthony G Sena, Ruth Costello, Belay Birlie, David Culliford, Caroline O'Leary, Henry Morgan, Theresa Burkard, Prof Daniel Prieto-Alhambra, Patrick Ryan

Abstract

Background: There is uncertainty around whether to use unicompartmental knee replacement (UKR) or total knee replacement (TKR) for individuals with osteoarthritis confined to a single compartment of the knee. We aimed to emulate the design of the Total or Partial Knee Arthroplasty Trial (TOPKAT) using routinely collected data to assess whether the efficacy results reported in the trial translate into effectiveness in routine practice, and to assess comparative safety.

Relative Risk of Cervical Neoplasms Among Copper and Levonorgestrel-Releasing Intrauterine System Users

Obstet Gynecol. 2020 Feb;135(2):319-327. doi: 10.1097/AOG.0000000000003656.

Authors

Matthew E Spotnitz, Karthik Natarajan, Patrick B Ryan, Carolyn L Westhoff

Abstract

Objective: To evaluate the relative risk of cervical neoplasms among copper intrauterine device (Cu IUD) and levonorgestrel-releasing intrauterine system (LNG-IUS) users.

Methods: We performed a retrospective cohort analysis of 10,674 patients who received IUDs at Columbia University Medical Center. Our data were transformed to a common data model and are part of the Observational Health Data Sciences and Informatics network. The cohort patients and outcomes were identified by a combination of procedure codes, condition codes, and medication exposures in billing and claims data. We adjusted for confounding with propensity score stratification and propensity score 1:1 matching.

Results: Before propensity score adjustment, the Cu IUD cohort included 8,274 patients and the

Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study

Lancet Rheumatol. 2020 Aug 21. doi: 10.1016/S2665-9913(20)30276-9. Online ahead of print.

Authors

Jennifer C E Lane, James Weaver, Kristin Kostka, Talita Duarte-Salles, Maria Tereza F Abrahao, Heba Alghoul, Osaid Alser, Thamir M Alshammari, Patricia Biedermann, Juan M Banda, Edward Burn, Paula Casajust, Mitchell M Conover, Aedin C Culhane, Alexander Davydov, Scott L DuVall, Dmitry Dymshyts, Sergio Fernandez-Bertolin, Kristina Fišter, Jill Hardin, Laura Hester, George Hripcsak, Benjamin Skov Kaas-Hansen, Seamus Kent, Sajan Khosla, Spyros Kolovos, Christophe G Lambert, Johan van der Lei, Kristine E Lynch, Rupa Makadia, Andrea V Margulis, Michael E Matheny, Paras Mehta, Daniel R Morales, Henry Morgan-Stewart, Mees Mosseveld, Danielle Newby, Fredrik Nyberg, Anna Ostropolets, Rae Woong Park, Albert Prats-Uribe, Gowtham A Rao, Christian Reich, Jenna Reps, Peter Rijnbeek, Selva Muthu Kumaran Sathappan, Martijn Schuemie, Sarah Seager, Anthony G Sena, Azza Shoaibi, Matthew Spotnitz, Marc A Suchard, Carmen O Torre, David Vizcaya, Haini Wen, Marcel de Wilde, Junqing Xie, Seng Chan You, Lin Zhang, Oleg Zhuk, Patrick Ryan, Daniel Prieto-Alhambra, OHDSI-COVID-19 consortium

Abstract

Renin-angiotensin system blockers and susceptibility to COVID-19: a multinational open science cohort study

medRxiv. 2020 Jun 12;2020.06.11.20125849. doi: 10.1101/2020.06.11.20125849. Preprint

Authors

Daniel R Morales, Mitchell M Conover, Seng Chan You, Nicole Pratt, Kristin Kostka, Talita Duarte Salles, Sergio Fernandez Bertolin, Maria Aragon, Scott L DuVall, Kristine Lynch, Thomas Falconer, Kees van Bochove, Cynthia Sung, Michael E Matheny, Christophe G Lambert, Fredrik Nyberg, Thamir M AlShammari, Andrew E Williams, Rae Woong Park, James Weaver, Anthony G Sena, Martijn J Schuemie, Peter R Rijnbeek, Ross D Williams, Jennifer C E Lane, Albert Prats Uribe, Lin Zhang, Carlos Areia, Harlan Krumholz, Daniel Prieto Alhambra, Patrick B Ryan, George Hripcsak, Marc A Suchard

Abstract

Introduction: Angiotensin converting enzyme inhibitors (ACEs) and angiotensin receptor blockers (ARBs) could influence infection risk of coronavirus disease (COVID-19). Observational studies to date lack pre-specification, transparency, rigorous ascertainment adjustment and international generalizability, with contradictory results.

Acute pancreatitis risk in type 2 diabetes patients treated with canagliflozin versus other antihyperglycemic agents: an observational claims database study

Curr Med Res Opin. 2020 Jul;36(7):1117-1124. doi: 10.1080/03007995.2020.1761312. Epub 2020 May 14.

Authors

Zhong Yuan, Frank DeFalco, Lu Wang, Laura Hester, James Weaver, Joel N Swerdel, Amy Freedman, Patrick Ryan, Martijn Schuemie, Rose Qiu, Jacqueline Yee, Gary Meininger, Jesse A Berlin, Norman Rosenthal

Abstract

Objective: Observational evidence suggests that patients with type 2 diabetes mellitus (T2DM) are at increased risk for acute pancreatitis (AP) versus those without T2DM. A small number of AP events were reported in clinical trials of the sodium glucose co-transporter 2 inhibitor canagliflozin, though no imbalances were observed between treatment groups. This observational study evaluated risk of AP among new users of canagliflozin compared with new users of six classes of other antihyperglycemic agents (AHAs).**Methods:** Three US claims databases were analyzed based on a

Diabetic ketoacidosis in patients with type 2 diabetes treated with sodium glucose co-transporter 2 inhibitors versus other antihyperglycemic agents: An observational study of four US administrative claims databases

Pharmacoepidemiol Drug Saf. 2019 Dec;28(12):1620-1628. doi: 10.1002/pds.4887. Epub 2019 Aug 27.

Authors

Lu Wang, Erica A Voss, James Weaver, Laura Hester, Zhong Yuan, Frank DeFalco, Martijn J Schuemie, Patrick B Ryan, Don Sun, Amy Freedman, Maria Alba, Joan Lind, Gary Meininger, Jesse A Berlin, Norman Rosenthal

Abstract

Purpose: To compare the incidence of diabetic ketoacidosis (DKA) among patients with type 2 diabetes mellitus (T2DM) who were new users of sodium glucose co-transporter 2 inhibitors (SGLT2i) versus other classes of antihyperglycemic agents (AHAs).

Methods: Patients were identified from four large US claims databases using broad (all T2DM

Prediction

Establishment and evaluation of a multicenter collaborative prediction model construction framework supporting model generalization and continuous improvement: A pilot study

Int J Med Inform. 2020 May 30;141:104173. doi: 10.1016/j.ijmedinf.2020.104173. Online ahead of print.

Authors

Yu Tian, Weiguo Chen, Tianshu Zhou, Jun Li, Kefeng Ding, Jingsong Li

Abstract

Background and objective: In recent years, an increasing number of clinical prediction models have been developed to serve clinical care. Establishing a data-driven prediction model based on large-scale electronic health record (EHR) data can provide a more empirical basis for clinical decision making. However, research on model generalization and continuous improvement is insufficiently focused, which also hinders the application and evaluation of prediction models in real clinical environments. Therefore, this study proposes a multicenter collaborative prediction model construction framework to build a prediction model with greater generalizability and continuous improvement capabilities while preserving patient data security and privacy.

Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation

BMC Med Res Methodol. 2020 May 6;20(1):102. doi: 10.1186/s12874-020-00991-3.

Authors

Jenna M Reps, Ross D Williams, Seng Chan You, Thomas Falconer, Evan Minty, Alison Callahan, Patrick B Ryan, Rae Woong Park, Hong-Seok Lim, Peter Rijnbeek

Abstract

Background: To demonstrate how the Observational Healthcare Data Science and Informatics (OHDSI) collaborative network and standardization can be utilized to scale-up external validation of patient-level prediction models by enabling validation across a large number of heterogeneous observational healthcare datasets.

Methods: Five previously published prognostic models (ATRIA, $CHADS_2$, $CHADS_2VASC$, Q-Stroke and Framingham) that predict future risk of stroke in patients with atrial fibrillation were

Development and validation of a prognostic model predicting symptomatic hemorrhagic transformation in acute ischemic stroke at scale in the OHDSI network

PLoS One. 2020 Jan 7;15(1):e0226718. doi: 10.1371/journal.pone.0226718. eCollection 2020.

Authors

Qiong Wang, Jenna M Reps, Kristin Feeney Kostka, Patrick B Ryan, Yuhui Zou, Erica A Voss, Peter R Rijnbeek, RuiJun Chen, Gowtham A Rao, Henry Morgan Stewart, Andrew E Williams, Ross D Williams, Mui Van Zandt, Thomas Falconer, Margarita Fernandez-Chas, Rohit Vashisht, Stephen R Pfohl, Nigam H Shah, Suranga N Kasthurirathne, Seng Chan You, Qing Jiang, Christian Reich, Yi Zhou

Abstract

Background and purpose: Hemorrhagic transformation (HT) after cerebral infarction is a complex and multifactorial phenomenon in the acute stage of ischemic stroke, and often results in a poor prognosis. Thus, identifying risk factors and making an early prediction of HT in acute cerebral infarction contributes not only to the selections of therapeutic regimen but also, more importantly, to the improvement of prognosis of acute cerebral infarction. The purpose of this study