Pharmacoepidemiology and Big Data Analytics: Challenges and Opportunities when Moving towards Precision Medicine

Andrea M. Burden*

Abstract: Pharmacoepidemiology is the study of the safety and effectiveness of medications following market approval. The increased availability and size of healthcare utilization databases allows for the study of rare adverse events, sub-group analyses, and long-term follow-up. These datasets are large, including thousands of patient records spanning multiple years of observation, and representative of real-world clinical practice. Thus, one of the main advantages is the possibility to study the real-world safety and effectiveness of medications in uncontrolled environments. Due to the large size (volume), structure (variety), and availability (velocity) of observational healthcare databases there is a large interest in the application of natural language processing and machine learning, including the development of novel models to detect drug–drug interactions, patient phenotypes, and outcome prediction. This report will provide an overview of the current challenges in pharmacoepidemiology and where machine learning applications may be useful for filling the gap.

Keywords: Big Data · Machine Learning · Medicine · Pharmacoepidemiology

Dr. Andrea Burden received her PhD in Pharmaceutical Sciences at the University of Toronto in 2014, where her doctoral dissertation specialized on methods and bias when using real world data in pharmacoepidemiology. In 2015, through the support of a Canadian Institutes for Health Research (CIHR) fellowship, she moved to the Netherlands to conduct a 3-year post-doctoral fellowship at the University of Maastricht. During this time, Dr. Burden was the co-lead of the pharmacoepidemiology research program at the Maastricht University Medical Centre with Prof. Dr. Frank de Vries, and she co-authored over 20 peer-reviewed manuscripts. In 2018, Dr. Burden was appointed as an assistant professor (tenure-track) at the ETH Zurich Institute for Pharmaceutical Sciences in the Department of Chemistry and Applied Biosciences. In this position she is the first chair of pharmacoepidemiology at the ETH Zurich. At the ETH Zurich, her group uses real-world patient data to conduct studies on the utilization, safety, and effectiveness of medications in everyday clinical practice. While her research interests are not focused on a specific clinical area, the majority of research in Dr. Burden's group is focused on chronic disease management within musculoskeletal and metabolic diseases.

1. Introduction

Pharmacoepidemiology is the study of the use and effects of medications post-market approval. In the modern era, medications are approved following rigorous clinical testing.[1] While large phase III Randomized Clinical Trials (RCTs) are the hallmark of drug approval, and provide the highest level of evidence regarding a causal assessment of the efficacy of a medication, it is well established that they cannot provide all the necessary information regarding medication safety and effectiveness.[2–4] This is due to the inherent limitations of RCTs, which are conducted in restricted patient groups (both size and representativeness), under highly controlled environments, and for relatively short durations. Importantly, RCTs are limited in the ability to detect rare adverse events and long-term effects. However, despite the limitations of RCTs they remain the 'gold standard' for determining the efficacy of a medication and the bench-mark for all observational studies.

Nevertheless, while RCTs provide optimal information on the short-term safety and efficacy of a medication under ideal controlled circumstances, the real-world safety and effectiveness remains largely unknown at the time of market approval. Consequently, post-market surveillance is essential to fill the gaps on the use, safety, and effectiveness of medications when used in more diverse patient populations and without strict monitoring of use.[5] This includes studying the real-world safety and effectiveness of medications in patients that were excluded from RCTs, and in every-day practice where prescribing occurs without randomization and medication adherence is not monitored.

There are a number of study designs available in observational post-market surveillance research, each with advantages and disadvantages.[6] Moving up the pyramid of evidence, these include case-reports, case-series, cross-sectional, time-series, case-control and cohort studies. Like any scientific research, the choice of study design for an observational pharmacoepidemiology study depends on the research question and data source. The most frequently used study design in pharmacoepidemiology studies is the cohort study design. Conceptually, these are the closest to RCTs, as patients are 'recruited' into the study based on the exposure and followed forward in time to observe the frequency of outcomes. However, rather than the researcher allocating patients to specific treatment groups, cohort studies observe ongoing medical care in observational data. The data sources for pharmacoepidemiology studies derive from routinely collected real-world healthcare data (RWD).[7,8] These databases allow for the inclusion of very large and representative patient populations that contain more complex patient groups (*e.g.* elderly patients).

Correspondence: Prof. A. M. Burden, E-mail: andrea.burden@pharma.ethz.ch
ETH Zurich, Institute for Pharmaceutical Sciences, Department of Chemistry and Applied Biosciences, HCI H 407, Vladimir-Prelog-Weg 4/10, CH-8093 Zurich

As a result, a well-conducted and analyzed pharmacoepidemiology study can provide information on the effectiveness and safety (particularly on rare adverse events) that is better than RCTs.[9] However, the observational nature of data collection brings about numerous challenges that must be considered when designing, analyzing and assessing a study to minimize bias and ensure the validity and accuracy of the results.

The remainder of this article reviews the databases available for pharmacoepidemiology research, the challenges that remain in current practice, and the potential for machine-learning analysis to provide new insights into medication safety and effectiveness research.

2. Real-World 'Big Data' in Pharmacoepidemiology

The field of pharmacoepidemiology has grown substantially in the past two decades as the availability and size of RWD has expanded. Unlike primary data collection, RWD include automated electronic records of interactions with healthcare services that are routinely collected for administration or payment purposes. These data are therefore not collected for research purposes and do not include systematic data collection that are tailored to the research question.[8] However, these secondary sources of data have a number of advantages as they are less time-consuming, cheaper, more representative of the general population, and large in size permitting the study of rare events and sub-group analyses.

Typically, RWD data within healthcare fall within two primary categories – electronic medical records or administrative claims data.[7] While the former often provides clinical information regarding symptoms and diagnostic test results, the later provides large and representative information at the population level. Importantly, both forms of RWD include patient populations that may have been excluded from RCTs; such as the elderly, those in nursing homes, those with high comorbidities/polypharmacy, and pregnant women.

Traditionally, RWD include information on medications, primary care and specialist visits, hospitalizations and vital statistic registers. However, recently there has been an emergence of new data sources, including registries of rare disease, expensive drugs and medical devices; biobanks with biomedical and genetic data; and patient-generated data resulting from wearable devices, social media, or patient-reported outcome measures (PROMs). Linkage of these individual data sources over time provides an extremely large and rich data source for conducting pharmacoepidemiologic research (Fig. 1). Data linkage is extremely important in order to generate a complete picture of a patients' healthcare trajectory and permit an assessment of causality (*i.e.* medication exposure prior to outcome occurrence).

Data sources such as the UK Clinical Practice Research Datalink (CPRD), the Danish National Healthcare Register, and the Ontario healthcare data available through the Institute for Clinical Evaluative Sciences (ICES) are commonly used primary care and administrative data in pharmacoepidemiologic research.[10–19] These data sources can link patient information throughout the healthcare system using individual patient identifiers, thereby providing a near complete overview of a patient trajectory.

Once linked, these data sources include tens of tens-of-thousands, if not millions, of patient records with multiple years of follow-up stemming from different data sources. While the majority of the data are structured or semi-structured in nature, we are increasingly seeing the emergence of unstructured text-heavy data. Importantly, due the routine collection of data for monitoring of patient care or budgeting, the data are available quickly and at a relatively low cost. Thus, RWD in pharmacoepidemiology fulfills the '3 V's' of 'Big Data' – volume, variety and velocity.[20]

Over the previous decade, there have been further efforts to grow the size and completeness of RWD to permit powerful analysis within and across countries. This includes projects like the OHDSI platform in Europe,[21] the Canadian Institute for Health Research CNODES in Canada,[22] and the Sentinel Initiative in the US.[23] The projects develop common data models (CDMs) to permit fast and efficient epidemiologic and pharmacoepidemiologic studies across multiple datasets, thereby developing an extremely large body of evidence that is representative and reproducible across databases and countries.

3. Challenges in Pharmacoepidemiology

While RWD are large and rich sources of healthcare data, the observational nature brings a number of unique challenges.[8,24,25] As previously identified, the data were not collected for research purposes, which leads to concerns regarding the accuracy and validity of the data when used on pharmacoepidemiology studies. In particular, there may be concerns regarding how medications are dispensed in clinical practice, confounding, and missing data.

In contrast to RCTs where patients are randomly allocated to treatment groups, medical doctors base the selection of treatment on a number of factors including patient characteristics, severity of illness, available evidence and patient preference.[26] As a result, there may be important differences between patients receiving different medications. This imbalance between user groups can significantly bias study results if not properly controlled for in the study design or analysis. For example, if all patients with very severe illness receive medication B, while healthier patients receive medications A and C, it would not be surprising to find that patients with medication B have worse outcomes when compared

Fig. 1. Overview of centrally linked real-world data (RWD).
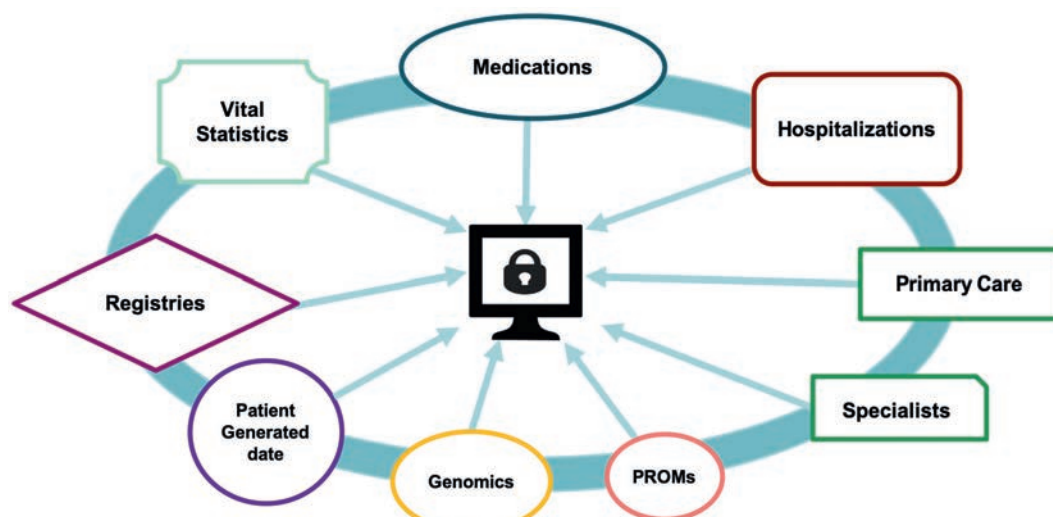
to medications A and C. This confounding by indication is one of the most challenging concepts in pharmacoepidemiology,[26,27] as the selection of an adequate control group is imperative to minimize the potential for selection bias leading to erroneous conclusions.

The propensity score (PS) analysis has become one of the most commonly used methods to control for confounding by indication. In brief, the PS is a summary statistic stemming from multilevel modelling to combine all confounding factors into a single score that can be used in adjustment, matching or stratification to balance a patients' likelihood for receiving a medication.[28] While a PS analysis typically uses only observed variables that are identified *a priori* as confounders (*i.e.* variables that are associated with both the exposure and outcome, but do not lie on the causal pathway), the high-dimensional PS (HDPS) is an automated data-driven approach that includes a large number of variables that may serve as proxies for unmeasured variables.[29] However, implementing HDPS analysis in time-varying models is challenging, and critics often cite the 'black-box' effect of variable selection that fails to provide transparent estimates and replication is difficult.[30]

While RWD bring about a number of methodological and analytical challenges, they have led to important advances in clinical practice and public health policy. Indeed, evidence produced through pharmacoepidemiology studies is often at the forefront of changes to clinical guideline recommendations and regulatory decision-making. Moreover, due to the large size of RWD, it is often powered to conduct sub-group analyses. This is a critical advantage, particularly as the field of medicine aims to move towards precision medicine.

Indeed, one of the primary goals in clinical pharmacoepidemiology research is to identify the optimal risk-benefit profile for a medication to guide physician prescribing.[31] However, in highly heterogeneous patient populations, this can be challenging when including the entire patient population as a whole. Rather, targeted sub-group analyses based on clinical factors, or novel clustering analysis, may reveal differential treatment response profiles. As such, this approach would be a first step towards tailored (or precision) medicine (Fig. 2).

However, while traditional methods such as survival analysis and logistic regression, have provided significant advancements in our knowledge of the utilization, safety and (comparative) effectiveness of medications, there remains a gap in our understanding of complex patient phenotypes, polypharmacy and dynamic longitudinal progression of chronic diseases. This is where machine-learning methods may have a substantial role within pharmacoepidemiology. In particular, machine learning methods may be superior in finding new features or non-linear relationships in the data, which can ultimately aide in hypothesis generation and drive research forward to improve treatment prediction.

4. Opportunities and Challenges for Machine-Learning in Pharmacoepidemiology

While machine-learning methods have seen widespread use within the financial and retail sectors, the uptake within healthcare, and particularly pharmacoepidemiology, has been limited. Most recent applications include using data-mining techniques to detect new safety signals in pharmacovigilance data,[32,33] prediction modeling,[34,35] and cluster analysis.[36,37] However, while there is increasing optimism of the future of machine learning in pharmacoepidemiology, it is important to consider the areas where machine learning would provide benefits over traditional methods. Currently, two burgeoning areas for machine-learning applications are in advancing our understanding of heterogeneous patient populations and improving the prediction of treatment response/safety.

Likely one of the most promising opportunities for machine-learning within pharmacoepidemiology is improving treatment prediction within heterogeneous patient groups. Indeed the high-dimensional nature of RWD lends itself well to developing pattern-recognition in heterogeneous patient groups, which can be used to improve patient characterization and longitudinal treatment prediction. In light of the trend towards precision medicine, one exciting opportunity for data-driven analyses lies in the application of deep-learning and natural language processing. Previous studies have applied unsupervised machine-learning methods to identify novel distinct phenotype clusters in heterogeneous patient populations, including applications in type 2 diabetes, chronic obstructive pulmonary disease, corticosteroid users and intensive care patients.[36–40] While there are a number of unsupervised and supervised clustering methods, the self-organizing map (SOM) is one of the most popular data-driven approaches due to the typology preserving mapping.[41] SOMs convert multidimensional data into a 2-dimensional visualization (Fig. 3), thereby allowing the identification of homogeneous sub-groups within data. Thus, while traditional stratification may be done based on a limited number of known clinical factors, such as age and sex, SOMs provide the opportunity to identify previously unknown clusters of common clinical characteristics within a diverse patient population. Neural network models are also frequently used for disease prediction within healthcare.[42,43]

4.1 *Application in Pharmacoepidemiology: Predicting Treatment Response in Rheumatoid Arthritis Patients*

Rheumatoid arthritis is a chronic and progressive autoimmune disease that is highly heterogeneous in the clinical presentation and progression.[44,45] While there is no cure for rheumatoid arthritis, there are a number of treatment options available to manage disease activity and optimally achieve clinical remission, including biologic disease modifying anti-rheumatic drugs (bDMARDs), namely the anti-tumor necrosis factor-alpha inhibitors (TNFi).
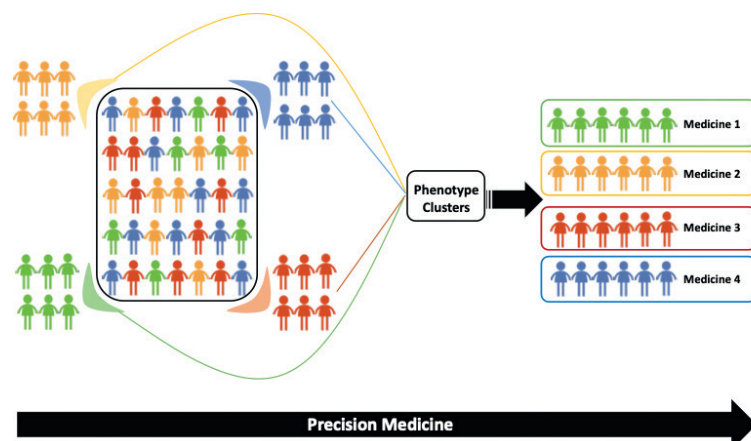

Fig. 2. Illustration of trajectory of precision medicine in pharmaco-epidemiology. Identifying homogeneous phenotype clusters from a heterogeneous patient population to improve treatment prediction.

Fig. 3. Simplified depiction of self-organizing map. The first-frame (1) identifies the visible input vectors and the visible output nodes. The second frame (2) visualizes the identification of the best matching unit (BMU). For each node, the algorithm finds the node closest to it, and the node with the closest distance is the BMU. This process is repeated for the different rows to identify the BMU. Frame three (3) is a visual depiction of a resulting SOM.
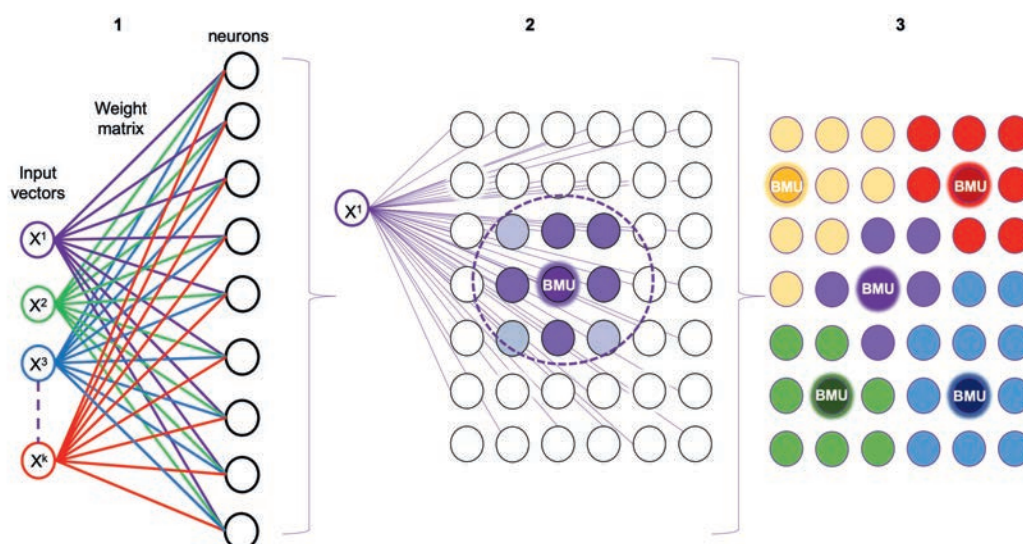
However, up to 40% of patients do not achieve clinical response (treatment failure) to the initial TNFi treatment,[46–52] while others (~20%) will lose response over time.[53] Thus, a primary goal in rheumatology is to develop personalized approaches to treatment selection to minimize the likelihood of treatment non-response, which delays remission thereby prolonging painful joint swelling and eventual bone erosions.

Despite a number of studies, the prediction of treatment-response remains sub-optimal in rheumatoid arthritis, and clinicians often adopt a trial-and-error approach to treatment selection. Although machine-learning algorithms are well-suited to develop prediction models that can handle complex time-dependent interactions of clinical factors, and are therefore an optimal approach to handle the heterogeneity in rheumatoid arthritis patients, to date they have not been applied to predict individual treatment success (*e.g.* comparative effectiveness). Norgeot *et al.*[42] recently concluded that use of recursive neural networks could be applied to predict changes in clinical disease activity index (CDAI) scores based on a combination of disease activity history, laboratory values and medications. In their analysis, Norgeot and colleagues used electronic medical data from two hospital settings to train and test their model performance to accurately predict changes in CDAI scores between visits.[42]

While recurrent neural networks are one approach, there is a need to develop models that are clinically explainable to ensure clinical application. If we develop a model that can predict treatment response to adalimumab (a TNFi) with 99% accuracy, but this is difficult to interpret or contains too many variables to have clinical relevance, it will certainly not improve patient care. Thus, while neural networks may increase prediction accuracy, they may limit our interpretability.[43] Conversely, logistic regression has a high-level of interpretability but low accuracy for complex longitudinal outcomes. Thus, within pharmacoepidemiology applications, we must search for the optimal balance between these two aspects.[54]

With this in mind, either LASSO regression or gradient boosting may be optimal methods to identify the most important patient-specific features for success with individual treatment options (*e.g.* adalimumab vs. abatacept). Given the high number of variables within RWD, the penalized LASSO (least absolute shrinkage and selection operator) regression, which builds the decision process based on only the most important clinical features, may be best suited for rare outcomes. Lund and colleagues recently developed a prediction model for 5-year mortality among elderly patients in the United States using LASSO regression.[34] Within rheumatology, only a small number of studies have applied LASSO regression to the prediction of rheumatoid arthritis

or treatment outcomes. Recently, Kan *et al.*[55] applied LASSO regression to successfully identify factors associated with unfilled prescriptions in a rheumatoid arthritis patient population. Thus, based on the simplicity and interpretability of LASSO regression, it is proposed that this may be an optimal approach to develop clinical tools to guide clinical decision making.

4.2 *Limitations of Machine-learning in Pharmacoepidemiology*

Machine-learning and artificial intelligence are currently buzzwords within healthcare. While they are highly attractive due to their ability to analyze diverse data types and incorporate multiple variables into risk stratifications and outcome prediction, there is a number of limitations, or areas of caution, that must be addressed. An important factor to consider when using RWD is that researchers must have sound knowledge of the data quality/structure, an awareness of the machine learning tools, and an understanding of the potential clinical implications. Not all research questions and data are suitable for machine-learning methods, and a blind application will lead to inappropriate use and potentially misleading research findings. With access to tens-of-thousands or even millions of patient records, it is highly important that care is taken to ensure appropriate use. Applying advanced algorithms universally to different data sources may lead to highly inaccurate results. Indeed, understanding how the data were built and the potential for missing or inaccurate data is vital to develop clinically relevant results. We must also be cognizant of the fact that patients and treatment practices vary across care settings, geographical regions and over time. It is therefore critical that predictive machine-learning algorithms in pharmacoepidemiology consider the robustness of algorithm in different settings (databases and time periods). Thus, it is highly encouraged to develop interdisciplinary teams that include pharmacoepidemiologists, data-scientists, and clinicians to ensure adequate knowledge on all aspects of databases limitations, algorithm appropriateness and clinical relevance are incorporated.

Finally, the possibility to improve treatment/outcome prediction in complex patient groups is one of the most exciting advantages with machine-learning, however it is important that we remain mindful that many prediction algorithms in machine learning research have been built and evaluated on metrics that lack clinical relevance. For example, a model with near perfect area under the receiver operating characteristic curve (AUC) may lack the needed sensitivity or positive predictive value for clinical use. Thus, as identified above there will likely be a trade-off between accuracy and interpretability. While LASSO regression is discussed in the above case example, gradient-boosting or ran-

dom-forest analyses may also be ideal approaches for application with clinical practice. Another aspect to consider is the potential to supplement prediction models with improved data visualization techniques to improve effective communication of results. Indeed, methods that can improve the understanding behind the predictions will be incredible useful within clinical practice and assist with transparency and reproducibility. For example, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) may be useful tools to provide insights into the direction each feature contributes within the model.

5. Conclusion

While machine-learning in pharmacoepidemiology is still in its infancy, there are a number of reasons to be excited about the opportunities of machine-learning to improve analytic efficiency, cluster identification and prediction estimates in clinical practice. Indeed, studies have demonstrated the promise of deep-learning approaches to uncover novel phenotype clusters and improve prediction in heterogeneous patient groups in RWD.[36,38,56,57] There are numerous clinical areas where prediction of treatment success or adverse events remains challenging due to the complex nature of the disease, or interactions over time with other comorbidities and medication use. As discussed here, complex patient groups, such as rheumatoid arthritis, is one area where machine-learning may have the greatest impact. Thus, it is no longer a question of can machine-learning be applied to the field of medicine and pharmacoepidemiology, but how do we ensure the appropriate use. Although the large scale and multidimensional nature of RWD is appealing to data-scientists, not all data, nor all research questions, will be suited to machine-learning methods. Future efforts must therefore be focused on identifying the areas where machine-learning provides advantages over traditional techniques. Moreover, given the clinical application of pharmacoepidemiology research, advancements in machine-learning with RWD must focus on transparency in methodology to ensure results are reproducible and translatable to clinical audiences. Nevertheless, while we must remain mindful of the limitations, the emergence of large RWD over the last two decades has opened the door for new discoveries in big data analytics, including machine-learning, and we have likely only scratched the surface of the possibilities.

Received: October 28, 2019

[1] B. L. Strom, in 'Pharmacoepidemiology, Fifth Edition', Eds. B. L. Strom, S. E. Kimmel, S. Hennessy, John Wiley & Sons, Ltd, **2012**, pp. 3.
[2] R. Lindsay, *Bone* **2007**, *40*, S32.
[3] W. A. Ray, *Am. J. Epidemiol.* **2003**, *158*, 915.
[4] J. Avorn, *N. Engl. J. Med.* **2007**, *357*, 2219.
[5] S. Schneeweiss, *Clin. Pharmacol. Ther.* **2016**, *99*, 262.
[6] B. L. Strom, in 'Pharmacoepidemiology, Fourth Edition', Ed. B. L. Strom, John Wiley & Sons, Ltd, Chichester, UK, **2007**, pp. 61.
[7] B. L. Strom, in 'Pharmacoepidemiology, Fifth Edition', Eds. B. L. Strom, S. E. Kimmel, S. Hennessy, John Wiley & Sons, Ltd, **2012**, pp. 158.
[8] S. Schneeweiss, J. Avorn, *J. Clin. Epidemiol.* **2005**, *58*, 323.
[9] 'From Randomized Controlled Trials to Observational Studies', S. L. Silverman, **2009**, DOI: 10.1016/j.amjmed.2008.09.030.
[10] J. H. M. Driessen, R. M. A. Henry, H. A. W. van Onzenoort, A. Lalmohamed, A. M. Burden, D. Prieto-Alhambra, C. Neef, H. G. M. Leufkens, F. de Vries, *Calcif. Tissue Int.* **2015**, *97*, 104.
[11] J. van Dalem, M. C. G. J. Brouwers, C. D. A. Stehouwer, A. Krings, H. G. M. Leufkens, J. H. M. Driessen, F. de Vries, A. M. Burden, M. C. G J Brouwers, C. D. A Stehouwer, A. Krings, H. G. M Leufkens, J. H. M Driessen, F. de Vries, A. M. Burden, *BMJ* **2016**, *354*, i3625.
[12] A. Lalmohamed, A. J. MacGregor, F. de Vries, H. G. M. Leufkens, T. P. van Staa, *PLoS One* **2013**, *8*, e65891.
[13] J. Van Dalem, M. C. G. J. Brouwers, C. D. A. Stehouwer, A. Krings, H. G. M. Leufkens, J. H. M. Driessen, F. De Vries, A. M. Burden, *BMJ* **2016**, *354*, i3625.
[14] E. Herrett, A. M. Gallagher, K. Bhaskaran, H. Forbes, R. Mathur, T. van Staa, L. Smeeth, *Int. J. Epidemiol.* **2015**, *44*, 827.
[15] M. A. Amiche, S. Abtahi, J. H. M. Driessen, P. Vestergaard, F. de Vries, S. M. Cadarette, A. M. Burden, *Arch. Osteoporos.* **2018**, *13*, 30.
[16] A. Lalmohamed, T. P. van Staa, P. Vestergaard, H. G. M. Leufkens, A. de Boer, P. Emans, C. Cooper, F. de Vries, *Am. J. Epidemiol.* **2016**, *184*, 58.
[17] A. M. Burden, J. M. Paterson, D. H. Solomon, M. Mamdani, D. N. Juurlink, S. M. Cadarette, *Osteoporos. Int.* **2012**, *23*, 1075.
[18] M. Tadrous, R. Matta, S. Greaves, S. Herschorn, M. M. Mamdani, D. N. Juurlink, T. Gomes, *JAMA Intern. Med.* **2019**, *179*, 1436.
[19] M. Tadrous, W. Khuu, G. Lebovic, M. B. Stanbrook, D. Martins, J. M. Paterson, M. M. Mamdani, D. N. Juurlink, T. Gomes, *Ann. Allergy. Asthma Immunol.* **2018**, *120*, 59.
[20] S. J. Mooney, D. J. Westreich, A. M. El-Sayed, *Epidemiology* **2015**, *26*, 390.
[21] G. Hripcsak, J. D. Duke, N. H. Shah, C. G. Reich, V. Huser, M. J. Schuemie, M. A. Suchard, R. W. Park, I. C. K. Wong, P. R. Rijnbeek, J. Van Der Lei, N. Pratt, G. N. Norén, Y. C. Li, P. E. Stang, D. Madigan, P. B. Ryan, *Stud. Health Technol. Inform.* **2015**, *216*, 574.
[22] I. S. Sketris, N. Carter, R. L. Traynor, D. Watts, K. Kelly, following contributing members of the CNODES Knowledge Translation Team: P. Ernst, J. M. Gamble, B. Hemmelgarn, C. Metge, M. Paterson, R. Platt, G. Teare, *Pharmacoepidemiol. Drug Saf.* **2019**, doi: 10.1002/pds.4738.
[23] R. Platt, J. S. Brown, M. Robb, M. McClellan, R. Ball, M. D. Nguyen, R. E. Sherman, *N. Engl. J. Med.* **2018**, *379*, 2091.
[24] J. A. Cramer, S. L. Silverman, D. T. Gold, *Curr. Med. Res. Opin.* **2007**, *23*, 2369.
[25] B. L. Strom, *Pharmacoepidemiol. Drug Saf.* **2001**, *10*, 389.
[26] J. L. F. Bosco, R. A. Silliman, S. S. Thwin, A. M. Geiger, D. S. M. Buist, M. N. Prout, M. U. Yood, R. Haque, T. Wei, T. L. Lash, *J. Clin. Epidemiol.* **2010**, *63*, 64.
[27] A. D. McMahon, *Pharmacoepidemiol. Drug Saf.* **2003**, *12*, 551.
[28] S. R. Johnson, G. A. Tomlinson, G. A. Hawker, J. T. Granton, B. M. Feldman, *Rheum. Dis. Clin. North Am.* **2018**, *44*, 203.
[29] S. Schneeweiss, J. A. Rassen, R. J. Glynn, J. Avorn, H. Mogun, M. A. Brookhart, *Epidemiology* **2009**, *20*, 512.
[30] R. Neugebauer, J. A. Schmittdiel, Z. Zhu, J. A. Rassen, J. D. Seeger, S. Schneeweiss, *Stat. Med.* **2015**, *34*, 753.
[31] G. S. Ginsburg, K. A. Phillips, *Health Aff. (Millwood)* **2018**, *37*, 694.
[32] H. Gurulingappa, L. Toldo, A. M. Rajput, J. A. Kors, A. Taweel, Y. Tayrouz, *Pharmacoepidemiol. Drug Saf.* **2013**, *22*, 1189.
[33] S. R. Feldman, *Semin. Arthritis Rheum.* **2015**, *44*, S16.
[34] J. L. Lund, T.-M. Kuo, M. A. Brookhart, A.-M. Meyer, A. F. Dalton, C. E. Kistler, S. B. Wheeler, C. L. Lewis, *Pharmacoepidemiol. Drug Saf.* **2019**, *28*, 584.
[35] S. S. Zhao, C. Hong, T. Cai, C. Xu, J. Huang, J. Ermann, N. J. Goodson, D. H. Solomon, T. Cai, K. P. Liao, *Rheumatology (Oxford)* **2019**, doi: 10.1093/rheumatology/kez375.
[36] L. E. G. W. Vanfleteren, M. A. Spruit, M. Groenen, S. Gaffron, V. P. M. van Empel, P. L. B. Bruijnzeel, E. P. A. Rutten, J. Op 't Roodt, E. F. M. Wouters, F. M. E. Franssen, *Am. J. Respir. Crit. Care Med.* **2013**, *187*, 728.
[37] K. C. Vranas, J. K. Jopling, T. E. Sweeney, M. C. Ramsey, A. S. Milstein, C. G. Slatore, G. J. Escobar, V. X. Liu, *Crit. Care Med.* **2017**, *45*, 1607.
[38] E. Ahlqvist, P. Storm, A. Käräjämäki, M. Martinell, M. Dorkhan, A. Carlsson, P. Vikman, R. B. Prasad, D. M. Aly, P. Almgren, Y. Wessman, N. Shaat, P. Spégel, H. Mulder, E. Lindholm, O. Melander, O. Hansson, U. Malmqvist, Å. Lernmark, K. Lahti, T. Forsén, T. Tuomi, A. H. Rosengren, L. Groop, *Lancet Diabetes Endocrinol.* **2018**, *6*, 361.
[39] A. Serra, P. Galdi, R. Tagliaferri, W*iley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1248.
[40] W. Wu, S. Bang, E. R. Bleecker, M. Castro, L. Denlinger, S. C. Erzurum, J. V Fahy, A. M. Fitzpatrick, B. M. Gaston, A. T. Hastie, E. Israel, N. N. Jarjour, B. D. Levy, D. T. Mauger, D. A. Meyers, W. C. Moore, M. Peters, B. R. Phillips, W. Phipatanakul, R. L. Sorkness, S. E. Wenzel, *Am. J. Respir. Crit. Care Med.* **2019**, *199*, 1358.
[41] T. Kohonen, *Biol. Cybern.* **1982**, *43*, 59.
[42] B. Norgeot, B. S. Glicksberg, L. Trupin, D. Lituiev, M. Gianfrancesco, B. Oskotsky, G. Schmajuk, J. Yazdany, A. J. Butte, *JAMA Netw. open* **2019**, *2*, e190606.
[43] F. Jaimes, J. Farbiarz, D. Alvarez, C. Martínez, *Crit. Care* **2005**, *9*, R150.
[44] J. A. Singh, K. G. Saag, S. L. Bridges, E. A. Akl, R. R. Bannuru, M. C. Sullivan, E. Vaysbrot, C. McNaughton, M. Osani, R. H. Shmerling, J. R. Curtis, D. E. Furst, D. Parks, A. Kavanaugh, J. O'Dell, C. King, A. Leong, E. L. Matteson, J. T. Schousboe, B. Drevlow, S. Ginsberg, J. Grober, E. W. St.Clair, E. Tindall, A. S. Miller, T. McAlindon, *Arthritis Care Res. (Hoboken)* **2016**, *68*, 1.
[45] J. S. Smolen, R. Landewé, J. Bijlsma, G. Burmester, K. Chatzidionysiou, M. Dougados, J. Nam, S. Ramiro, M. Voshaar, R. van Vollenhoven, D. Aletaha, M. Aringer, M. Boers, C. D. Buckley, F. Buttgereit, V. Bykerk, M. Cardiel, B. Combe, M. Cutolo, Y. van Eijk-Hustings, P. Emery, A. Finckh, C. Gabay, J. Gomez-Reino, L. Gossec, J.-E. Gottenberg, J. M. W. Hazes, T. Huizinga, M. Jani, D. Karateev, M. Kouloumas, T. Kvien, Z. Li, X. Mariette, I. McInnes, E. Mysler, P. Nash, K. Pavelka, G. Poór, C. Richez, P. van Riel, A. Rubbert-Roth, K. Saag, J. da Silva, T. Stamm, T. Takeuchi, R. Westhovens, M. de Wit, D. van der Heijde, *Ann. Rheum. Dis.* **2017**, *76*, 960.

[46] E. C. Keystone, A. F. Kavanaugh, J. T. Sharp, H. Tannenbaum, Y. Hua, L. S. Teoh, S. A. Fischkoff, E. K. Chartash, *Arthritis Rheum.* **2004**, *50*, 1400.

[47] E. C. Keystone, M. C. Genovese, L. Klareskog, E. C. Hsia, S. T. Hall, P. C. Miranda, J. Pazdur, S.-C. Bae, W. Palmer, J. Zrubek, M. Wiekowski, S. Visvanathan, Z. Wu, M. U. Rahman, GO-FORWARD Study, *Ann. Rheum. Dis.* **2009**, *68*, 789.

[48] J. S. Smolen, A. Beaulieu, A. Rubbert-Roth, C. Ramos-Remus, J. Rovensky, E. Alecock, T. Woodworth, R. Alten, *Lancet* **2008**, *371*, 987.

[49] L. Klareskog, D. van der Heijde, J. P. de Jager, A. Gough, J. Kalden, M. Malaise, E. M. Mola, K. Pavelka, J. Sany, L. Settas, J. Wajdula, R. Pedersen, S. Fatenejad, M. Sanda, *Lancet* **2004**, *363*, 675.

[50] P. E. Lipsky, *N. Engl. J. Med.* **2000**, *343*, 1594.

[51] J. M. Kremer, H. K. Genant, L. W. Moreland, A. S. Russell, P. Emery, C. Abud-Mendoza, J. Szechinski, T. Li, Z. Ge, J.-C. Becker, R. Westhovens, *Ann. Intern. Med.* **2006**, *144*, 865.

[52] G. R. Burmester, W. F. Rigby, R. F. van Vollenhoven, J. Kay, A. Rubbert-Roth, A. Kelman, S. Dimonaco, N. Mitchell, *Ann. Rheum. Dis.* **2016**, *75*, 1081.

[53] A. Rubbert-Roth, A. Finckh, *Arthritis Res. Ther.* **2009**, *11 Suppl 1*, S1.

[54] Y. Luo, H.-H. Tseng, S. Cui, L. Wei, R. K. Ten Haken, I. El Naqa, *BJR|Open* **2019**, *1*, 20190021.

[55] H. J. Kan, K. Dyagilev, P. Schulam, S. Saria, H. Kharrazi, D. Bodycombe, C. T. Molta, J. R. Curtis, *Arthritis Res. Ther.* **2018**, *20*, 79.

[56] T. Ahmad, L. H. Lund, P. Rao, R. Ghosh, P. Warier, B. Vaccaro, U. Dahlström, C. M. O'Connor, G. Michael Felker, N. R. Desai, *J. Am. Heart Assoc.* **2018**, *7*, 1.

[57] T. Karpati, M. Leventer-Roberts, B. Feldman, C. Cohen-Stavi, I. Raz, R. Balicer, *PLoS One* **2018**, *13*, e0207096.