Protocol: Deep Learning Comparison

Research area: methodological, disease epidemiology
Investigators: Chungsoo Kim, Luis H. John, Jenna M. Reps, Egill A. Fridgeirsson
Protocol version: 2023-08-21

Watch Out!

This study requires a recent Nvidia graphics card to execute. Please refer to the *Hardware Requirements* section for more details.

Lay summary

There are existing tools that can calculate a personalized estimate for the probability of developing a disease in the future. These tools are known as prediction models. Such models could be used to aid doctors when making medical decisions. In recent years, new methods to calculate such estimates have been developed. However, these methods have mostly been used in other domains, for example on image data or text data. It is unknown how well these new methods work for estimating personalized risk on patient data.

This research thoroughly compares the accuracy of traditionally used methods with the new methods. A large study is conducted on databases across the world to ensure that the new methods work well not only on our patients, but for every patient.

We focus on three different diseases for which we calculate personalized estimates. Those are dementia in persons above the age of 55, lung cancer in persons above the age of 45 and bipolar disorder in persons misdiagnosed with major depressive disorder.[1]

Technical summary

This study aims to compare conventional and deep learning prediction algorithms to determine if the latter provides benefits for estimating clinical outcome risk. We predict three clinically relevant outcomes in three respective study populations: (1) dementia in persons above the age of 55; (2) lung cancer in persons above the age of 45; (3) bipolar disorder in persons misdiagnosed with major depressive disorder.

The analysis is conducted on a network of observational healthcare databases that have been converted to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Data sources that will be used to develop and validate clinical prediction models include Optum De-identified Electronic Health Record Dataset (Optum EHR), Optum De-identified Clinformatics Data Mart - Socio-Economic Status (Optum SES), Integrated Primary Care Information (IPCI) and Ajou University School of Medicine (AUSOM). Moreover, this study encourages external partners to participate in executing the analysis on their own data sources, which will ensure better generalizability of results.

This retrospective patient-level prediction study develops and evaluates a set of conventional and deep learning prediction algorithms including logistic regression, gradient boosting (XGBoost), residual neural network (ResNet), transformer neural network (FT-Transformer). To evaluate the performance we calculate the discrimination ability of the model using the area under the receiver operating characteristic curve (AUROC) and the model calibration using the e-statistic and calibration-in-the-large complemented by an inspection of calibration plots.

This study benefits public health by presenting a set of clinical prediction models that can support the clinical decision-making process of a healthcare professional to predict one of the prognostic outcomes to improve patients' health.

Specific Aims, Objectives, and Rationale

Specific aims:

- Assess the added value of massive observational healthcare data for the development of deep learning models.
- Extend the standardized patient-level prediction framework to include deep learning algorithms

Objectives:

- Compare prediction performance on observational healthcare data for conventional and deep learning prediction methods.

Rationale:

- Deep learning models had a profound impact on the world, albeit limited to domains using primarily unstructured data such as imaging or natural language. This work studies the potential benefits of state-of-the-art deep learning prediction models on observational healthcare data over conventional prediction methods.

Study Background

Early diagnosis of individuals at high risk of disease allows for improved care and risk-factor targeted intervention. Prognostic models can provide such risk estimates and support a healthcare provider's clinical decision making which can ultimately improve patients' health.[2, 3]

Conventional prediction methods such as the general statistical model of logistic regression have been originally developed and popularized as early as 1944.[4] And to this day, logistic regression remains a state-of-the-art method to develop robust clinical prediction models, despite the impressive advances in more complex prediction approaches using deep learning.[5, 6] Similarly, tree-based models continue to outperform deep learning on typical tabular data.[7]

Deep learning models had a profound impact on the world, albeit limited to domains using primarily unstructured data such as imaging or natural language. Observational healthcare data presents itself primarily in the form of structured data, where conventional prediction methods such as logistic regression continue to achieve equal performance as compared to considerably more complex deep learning models.[8] Observational healthcare data was found to limit the efficacy of deep learning methods due to high sparsity, high dimensionality, and heterogeneity.[8-11]

However, recent work shifted the focus of deep learning also to structured data and new promising methods such as Tabnet have emerged.[12] This work presents a comprehensive comparison of conventional and deep learning methods to predict three clinically relevant health outcomes.

Background on dementia prediction

Dementia is an umbrella term to describe various illnesses that affect cognition and may lead to mental degradation.[13] All types of dementia are progressive, meaning that symptoms may be relatively mild at first but worsen with time, usually over the course of several years. Symptoms include problems with memory, thinking, problem-solving or language, changes in emotion, perception or behavior.[14]

Although getting older is the most significant risk factor for dementia, there exist measures that a patient can take to slow down dementia progression, such as physical activity, healthy eating, no smoking or drinking of alcohol, staying mentally and socially active.[14]

Therefore, a prediction model can be a useful tool to quantify the risk of a patient to develop dementia in the next years and take risk-factor-targeted intervention.

Background on lung cancer prediction

Lung cancer is the second most common cancer in the world and has one of the highest mortality rates among cancers. Respiratory cancer including tracheal, bronchus, and lung cancers increased by 23.3% from 2010 to 2019, with five-year survival rates (7% to 25%) significantly lower than other major cancers.[15-16]

The only recommended screening method for lung cancer is low-dose computed tomography.[17] This method has the limitation that it can only be applied to patients who meet predefined screening criteria. However, for those who do not meet these screening criteria, it may actually delay the detection of lung cancer, so there is a need for a predictive model that can identify patients at high risk of lung cancer.[18]

In a previous study, a machine learning model using routinely collected data outperformed the widely used mPLCOm2012 model, suggesting that machine learning methods can be used to predict lung cancer risk.[19]

Still few studies have applied deep learning to develop predictive models and compare their performance.

Background on bipolar disorder prediction

Bipolar disorder is a mental health condition that causes extreme mood swings between emotional highs (such as mania or hypomania) and lows (depression). The diagnosis of early bipolar disorder is often confused with diagnoses such as major depressive disorder because of the common symptom of depression. In fact, about 0.8 to 3.9% of patients diagnosed with major depressive disorder change their diagnosis to bipolar disorder during follow-up.

Bipolar disorder has a poorer prognosis than major depressive disorder and requires a different treatment strategy. In particular, even common treatments for depressive symptoms, such as the use of antidepressants, are controversial to use in people with bipolar disorder. In some bipolar disorders, antidepressant use can worsen symptoms, so being misdiagnosed with major depressive disorder can worsen a patient's clinical course.[20]

Study type and design

This is a retrospective patient-level prediction study.

Methodological: This study aims to develop and compare conventional and deep learning patient-level prediction models using machine learning best practices.

Disease epidemiology: This study develops patient-level prediction models for three clinically relevant health outcomes.

Source of Data

This study used observational data from administrative claims and electronic health records that were mapped into the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).[21] The OMOP CDM provides a standardized data structure and vocabulary, which enables computer-executed analysis to be shared among researchers and facilities.

At the point various data sources have already been confirmed to participate in this network study:

Database	Handle	Data type	Country
Optum De-identified Electronic Health Record Dataset	Optum EHR	EHR	US

Optum De-identified Clinformatics Data Mart - Socio-Economic Status	Optum SES	Claims	US
Integrated Primary Care Information	IPCI	GP	Netherlands
Ajou University School of Medicine	AUSOM	EHR	South Korea

Optum De-identified Electronic Health Record Dataset (Optum EHR)

Optum's longitudinal EHR repository is derived from dozens of healthcare provider organizations in the United States, that include more than 57 contributing sources and 111K sites of care: treating more than 106 million patients receiving care in the United States. The data is certified as de-identified by an independent statistical expert following HIPAA statistical de-identification rules and managed according to Optum® customer data use agreements . Clinical, claims and other medical administrative data is obtained from both Inpatient and Ambulatory electronic health records (EHRs), practice management systems and numerous other internal systems. Information is processed, normalized, and standardized across the continuum of care from both acute inpatient stays and outpatient visits. Optum® data elements include demographics, medications prescribed and administered, immunizations, allergies, lab results (including microbiology), vital signs and other observable measurements, clinical and inpatient stay administrative data and coded diagnoses and procedures. In addition, Optum® uses natural language processing (NLP) computing technology to transform critical facts from physician notes into usable datasets. The NLP data provides detailed information regarding signs and symptoms, family history, disease related scores (i.e. RAPID3 for RA, or CHADS2 for stroke risk), genetic testing, medication changes, and physician rationale behind prescribing decisions that might never be recorded in the EHR.

Optum De-identified Clinformatics Data Mart - Socio-Economic Status (Optum SES)

Optum's Clinformatics® Data Mart (CDM) is derived from a database of administrative health claims for members of large commercial and Medicare Advantage health plans. The database includes approximately 17-19 million annual covered lives, for a total of over 65 million unique lives over a 14-year period (1/2007 through 12/2021). Clinformatics® Data Mart is statistically de-identified under the Expert Determination method consistent with HIPAA and managed according to Optum® customer data use agreements. CDM administrative claims submitted for payment by providers and pharmacies are verified, adjudicated and de-identified prior to inclusion. This data, including patient-level enrollment information, is derived from claims submitted for all medical and pharmacy health care services with information related to health care costs and resource utilization. The population is geographically diverse, spanning all 50 states.

Integrated Primary Care Information (IPCI)

The Integrated Primary Care Information (IPCI) database is a longitudinal observational database containing routinely collected data from computer-based patient records of a selected group of GPs throughout the Netherlands. IPCI was started in 1992 by the department of Medical Informatics of the Erasmus University Medical Center in Rotterdam.

The current database includes patient records from 2006 on, when the size of the database started to increase significantly. In 2016, IPCI was certified as Regional Data Center. Since 2019 the data is also standardized to the OMOP CDM.[22]

AUSOM

The Ajou University School of Medicine (AUSOM) database is the EHR database of 2 752 765 patients who visited the Ajou University Medical Center from 1994 to 2023 Feb. Ajou University Medical Center in South Korea is a tertiary teaching hospital with 1,108 beds, 33 medical departments, and 23 operating rooms. The AUSOM database is in the form of the OMOP CDM version 5.3.

Feasibility count

A cohort diagnostics package will be provided, which can be executed prior to the start of the study. The cohort diagnostics package will provide statistics including feasibility counts.

Definition of the study populations and health outcomes

For this study, we develop various prediction models for three clinically relevant health prediction problems: (1) dementia in persons above the age of 55; (2) bipolar in patients newly diagnosed with major depressive disorder, (3) lung cancer in patients who are cancer free at their first outpatient visit in the year.

Dementia prediction

We included persons between 55-84 years of age with an index between 1 January 2014 – 31 December 2014. We use the latest recorded visit to a healthcare provider as the index event.

We exclude persons with prior dementia as defined by our outcome. Moreover, we exclude patients with disease records indicating subtypes of dementia including Parkinsonism (concept 4140090), Korsakoff's psychosis (concept 4288013), Huntington's chorea (concept 374341), human immunodeficiency virus infection (concept 439727, Creutzfeldt-Jakob disease (concept 372241) and all hierarchical descendants of these concepts according to the SNOMED medical terms hierarchy. In addition, we exclude patients with records of developmental mental disorder (concept 4043545), demyelinating disease of central nervous

system (concept 375801), and degenerative disease of the central nervous system (concept 4213310).

We exclude persons with a record of any drug included in the ATC code N06D of anti-dementia drugs (concept 21604792).

We exclude persons with a record of traumatic brain injury (concept 4132546), Traumatic AND/OR non-traumatic brain injury (concept 4133611), and lesion of brain (concept 4200516), and all hierarchical descendants of these concepts according to the SNOMED medical terms hierarchy.

We exclude persons with cognitive impairment including memory impairment (concept 4304008), impaired cognition (concept 443432), and mild cognitive disorder (concept 4297400), and all hierarchical descendants of these concepts according to the SNOMED medical terms hierarchy.

All exclusion criteria are assessed all time prior to the index date.

In addition, participants require 365 days of continuous observation time before the index date in which candidate predictors are assessed. This period is consistent with other models in literature that were developed on observational data and improves coverage of persons that may not be part of a database for a long time. Moreover, following the recommendations of an empirical analysis of dealing with patients who are lost to follow-up, we allow patients to leave the cohort at any time during the time-at-risk period. The time-at-risk period for patient ends after five years following the index date.[23]

We investigate the outcome of dementia for the first time in a person's history within 5 years following the index date.

Dementia is defined as Senile degeneration of brain (concept 373179), or Frontotemporal dementia (concept 4043378), or dementia (concept 4182210), and all hierarchical descendants of these concepts according to the SNOMED medical terms hierarchy.
Moreover, dementia is defined as senility (concept 435088), or organic mental disorder (concept 374009), or diffuse Lewy body disease (concept 380701), or cerebral degeneration associated with another disorder (concept 4104700), or amnestic disorder (concept 372608), or age-related cognitive decline (4009705).

We explicitly exclude from the definition of dementia senile and presenile organic psychotic conditions (concept 4152048), and postconcussion syndrome (concept 372610), and general paresis – neurosyphilis (concept 377788), and drug-induced dementia (concept 376095), and dementia following injury caused by exposure to ionizing radiation (concept 42535731), and dementia caused by volatile inhalant (concept 37311999), and dementia caused by toxin (concept 36717598), or dementia caused by heavy metal exposure (concept 37116464), and all hierarchical descendants of these concepts according to the SNOMED medical terms hierarchy.

Bipolar in MDD patients prediction

In Nestsiarovich *et al.* 2021 a simple score based model was developed to predict which patients who are newly diagnosed with MDD go on to develop bipolar within 1-year of their MDD diagnosis. We use the same prediction task in this study except we use later data (ensuring we use pre-covid data due to differences in healthcare treatment during the initial period of the pandemic).

Nestsiarovich, A., Reps, J.M., Matheny, M.E. *et al.* Predictors of diagnostic transition from major depressive disorder to bipolar disorder: a retrospective observational network study. *Transl Psychiatry* 11, 642 (2021). https://doi.org/10.1038/s41398-021-01760-6

Lung cancer prediction

In Chandran *et al.* 2023 a lung cancer prediction model was developed to predict 3-year risk of Lung cancer in patients with no prior cancer or cancer screening. We use the same cohorts except we use a later year for the outpatient visit (ensuring we use pre-covid data).

Chandran U, Reps J, Yang R, et al. Machine Learning and Real-World Data to Predict Lung Cancer Risk in Routine Care. Cancer Epidemiology, Biomarkers & Prevention : a Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology. 2023 Mar;32(3):337-343. DOI: 10.1158/1055-9965.epi-22-0873. PMID: 36576991; PMCID: PMC9986687.

Covariates

Variables cured automatically using OHDSI's FeatureExtraction package will be used as covariates in the prediction models.

In selecting covariates all demographics, diagnoses, drug use histories, and Charlson comorbidity index of the study population will be used as predictive covariates. As the criterion for extracting these covariates, the time point of one year in the past is applied, and all variables will be dichotomized. In other words, it will be used as 1 if the corresponding code exists within the past 1 year from the index date, and 0 if it does not exist.

Add fifty covariates in this setting

Statistical analysis

We used the patient-level prediction framework for model development and validation. This framework enables the development of analysis packages in R that can be shared across data sites mapped to the OMOP CDM.[3]

The prediction approach is presented in Figure 1. Among a population at risk, we predict which patients at a defined moment in time (the index) will experience some outcome during

a time-at-risk. Prediction is done using only information about the patients in an observation window prior to the index.

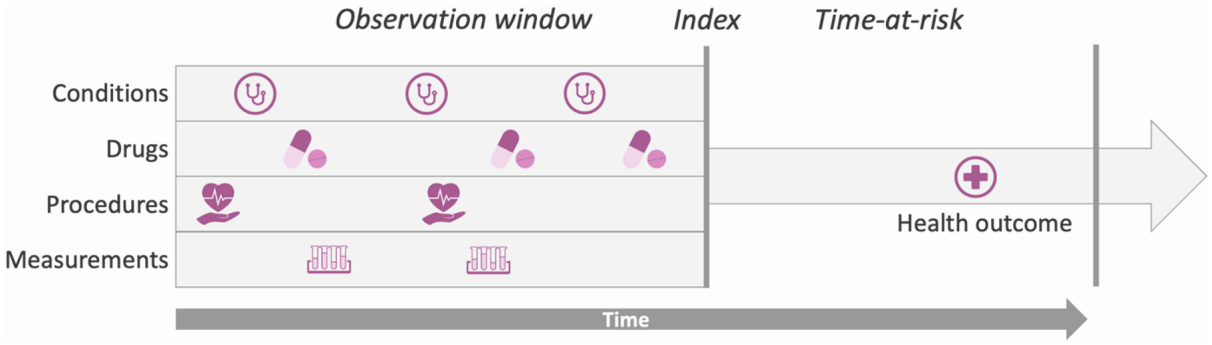

Figure 1. Prediction approach.

We used a train-test split by person to perform internal validation. Each person appeared only once in the datasets, because we only use their latest visit to a healthcare provider. In each development cohort, a random sample of 75% of patients was used to develop the prediction models and the remaining 25% of patients were used to internally validate the models.

The analysis constitutes a comparison of several prediction methods.

Method	**Notable design choice**
Logistic Regression	L1 regularization
Gradient Boosting Machine	
ResNet	Embedding layer and dropout
Transformer	

Our baseline model is LASSO regularized logistic regression with 3-fold cross validation to learn the optimal regularization hyperparameter through an adaptive search using Cyclops.[24] XGBoost is used to train the Gradient Boosting model. The architecture of the ResNet and FT-Transformer are detailed in their respective research paper (https://arxiv.org/pdf/2106.11959.pdf).

To evaluate the performance, we calculate the overall discrimination of the model using the area under the receiver operating characteristic curve (AUROC) and the model calibration (E-statistic and calibration plot). The AUROC indicated the probability that for two randomly selected patients, the patient who gets the outcome will be assigned a higher risk. The model calibration presented in a plot to examine agreement between predicted and observed risk across deciles of predicted risk. Calibration assessment is then performed visually and using the single value metric E-statistic to both provide a better impression of the direction and scale of miscalibration as well as compare calibration across models. Summary statistics are reported for the test samples.

Plans for addressing confounding

Confounding is not a big problem when performing prediction, because we are not concerned with identifying the exact effect of a variable on another (causation). For prediction, we are looking to find out what is the 'most likely' value of a dependent variable given a set of candidate predictors (association). Therefore, confounding will not be further addressed.

Plans for addressing missing data

Missing data in observational healthcare data is common, because such routinely collected data is not primarily intended for the purpose of research. It is likely that there exists a large portion of missing records due to a person experiencing symptoms, but not visiting a healthcare provider, leading to no record in a database. This is a known limitation of observational healthcare data and we do not intend to address it further.

The data in this study uses exclusively binary data that can take the value 0 (not recorded) and 1 (recorded). Missing data will by default take the value 0 and no imputations will be done.

This approach may not be appropriate for sine key data that relate to the respective health outcomes and will present a limitation.

Plans for disseminating and communicating study results

The final results of this study will be disseminated as a journal article.

All statistical analysis code will be made publicly available. This includes the study package for the network study, a cohort diagnostics package, and notable prediction models.

Hardware requirements

Due to higher complexity, deep learning models may require longer training times than conventional models on a CPU. Because of this, deep learning frameworks generally provide support for GPUs, which can speed up training significantly by taking advantage of their massive parallelization capabilities. For this study a CUDA enabled GPU as listed by NVIDIA is required (https://developer.nvidia.com/cuda-gpus). A large amount of at least 12 GB, but better 24 GB or more GPU memory is recommended. This allows for an increase of the batch size, which can further speed up execution times.

References

[1] Nestsiarovich, A., Reps, J.M., Matheny, M.E. *et al.* Predictors of diagnostic transition from major depressive disorder to bipolar disorder: a retrospective observational network study. *Transl Psychiatry* 11, 642 (2021). https://doi.org/10.1038/s41398-021-01760-6

[2] Todd A. Alonzo, Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating: By Ewout W. Steyerberg, *American Journal of Epidemiology*, Volume 170, Issue 4, 15 August 2009, Page 528, https://doi.org/10.1093/aje/kwp129

[3] Jenna M Reps, Martijn J Schuemie, Marc A Suchard, Patrick B Ryan, Peter R Rijnbeek, Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data, *Journal of the American Medical Informatics Association*, Volume 25, Issue 8, August 2018, Pages 969–975, https://doi.org/10.1093/jamia/ocy032

[4] Berkson, Joseph. “Application of the Logistic Function to Bio-Assay.” *Journal of the American Statistical Association* 39 (1944): 357-365.

[5] Cynthia Yang, Jan A Kors, Solomon Ioannou, Luis H John, Aniek F Markus, Alexandros Rekkas, Maria A J de Ridder, Tom M Seinen, Ross D Williams, Peter R Rijnbeek, Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review, *Journal of the American Medical Informatics Association*, Volume 29, Issue 5, May 2022, Pages 983–989, https://doi.org/10.1093/jamia/ocac002

[6] Suchard MA, Simpson SE, Zorych I, Ryan P, Madigan D. Massive parallelization of serial inference algorithms for a complex generalized linear model. ACM Trans Model Comput Simul. 2013 Jan;23(1):10.1145/2414416.2414791. doi: 10.1145/2414416.2414791. PMID: 25328363; PMCID: PMC4201181.

[7] L'eo Grinsztajn, Edouard Oyallon, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? arXiv pre-print server. 2022.

[8] Chen, D., Liu, S., Kingsbury, P. *et al.* Deep learning and alternative learning strategies for retrospective real-world clinical data. *npj Digit. Med.* 2, 43 (2019). https://doi.org/10.1038/s41746-019-0122-0

[9] Wang F, Casalino LP, Khullar D. Deep Learning in Medicine—Promise, Progress, and Challenges. *JAMA Intern Med.* 2019;179(3):293–294. doi:10.1001/jamainternmed.2018.7117

[10] Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc. 2018 Oct 1;25(10):1419-1428. doi: 10.1093/jamia/ocy068. PMID: 29893864; PMCID: PMC6188527.

[11] Deep learning for healthcare: review, opportunities and challenges." Briefings in bioinformatics 19.6 (2018): 1236-1246.

[12] Arik, S. Ö., & Pfister, T. (2021). TabNet: Attentive Interpretable Tabular Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(8), 6679-6687. https://doi.org/10.1609/aaai.v35i8.16826

[13] Stephan, B., Kurth, T., Matthews, F. *et al.* Dementia risk prediction in the population: are screening models accurate?. *Nat Rev Neurol* 6, 318–326 (2010). https://doi.org/10.1038/nrneurol.2010.54

[14] The progression, signs and stages of dementia 2023 [Available from: https://www.alzheimers.org.uk/about-dementia/symptoms-and-diagnosis/how-dementia-progresses/progression-stages-dementia.

[15] Global, regional, and national burden of respiratory tract cancers and associated risk factors from 1990 to 2019: a systematic analysis for the Global Burden of Disease Study 2019, Ebrahimi, Hedyeh et al. The Lancet Respiratory Medicine, Volume 9, Issue 9, 1030 - 1049

[16] Huang J, Deng Y, Tin MS, Lok V, Ngai CH, Zhang L, Lucero-Prisno DE 3rd, Xu W, Zheng ZJ, Elcarte E, Withers M, Wong MCS. Distribution, Risk Factors, and Temporal Trends for Lung Cancer Incidence and Mortality: A Global Analysis. Chest. 2022 Apr;161(4):1101-1111. doi: 10.1016/j.chest.2021.12.655. Epub 2022 Jan 11. PMID: 35026300.

[17] Jin J. Screening for Lung Cancer. *JAMA.* 2021;325(10):1016. doi:10.1001/jama.2021.1799

[18] Faselis C, Nations JA, Morgan CJ, et al. Assessment of Lung Cancer Risk Among Smokers for Whom Annual Screening Is Not Recommended. *JAMA Oncol.* 2022;8(10):1428–1437. doi:10.1001/jamaoncol.2022.2952

[19] Gould MK, Huang BZ, Tammemagi MC, Kinar Y, Shiff R. Machine Learning for Early Lung Cancer Identification Using Routine Clinical and Laboratory Data. Am J Respir Crit Care Med. 2021 Aug 15;204(4):445-453. doi: 10.1164/rccm.202007-2791OC. PMID: 33823116.

[20] Gitlin, M.J. Antidepressants in bipolar depression: an enduring controversy. *Int J Bipolar Disord* 6, 25 (2018). https://doi.org/10.1186/s40345-018-0133-9Gitlin, M.J. Antidepressants in bipolar depression: an enduring controversy. *Int J Bipolar Disord* 6, 25 (2018). https://doi.org/10.1186/s40345-018-0133-9

[21] J Marc Overhage, Patrick B Ryan, Christian G Reich, Abraham G Hartzema, Paul E Stang, Validation of a common data model for active safety surveillance research, *Journal of the American Medical Informatics Association*, Volume 19, Issue 1, January 2012, Pages 54–60, https://doi.org/10.1136/amiajnl-2011-000376

[22] Maria A J de Ridder, Marcel de Wilde, Christina de Ben, Armando R Leyba, Bartholomeus M T Mosseveld, Katia M C Verhamme, Johan van der Lei, Peter R Rijnbeek, Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands, *International Journal of Epidemiology*, Volume 51, Issue 6, December 2022, Pages e314–e323, https://doi.org/10.1093/ije/dyac026

[23] Walters, K., Hardoon, S., Petersen, I. *et al.* Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data. *BMC Med* 14, 6 (2016). https://doi.org/10.1186/s12916-016-0549-y

[24] Suchard MA, Simpson SE, Zorych I, Ryan P, Madigan D. Massive parallelization of serial inference algorithms for a complex generalized linear model. ACM Trans Model Comput Simul. 2013 Jan;23(1):10.1145/2414416.2414791. doi: 10.1145/2414416.2414791. PMID: 25328363; PMCID: PMC4201181.