medRxiv preprint doi: https://doi.org/10.1101/2019.12.26.19015909; this version posted December 30, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Accurate and reproducible prediction of ICU readmissions

Dinh-Phong NGUYEN[1, 2,✉], **Nicolas PARIS**[1], **and Adrien PARROT**[1]

[1]WIND-DSI, AP-HP, Paris, France
[2]Sorbonne Université, UPMC Univ Paris 06, Paris, France

Readmission in the intensive care unit (ICU) is associated with poor clinical outcomes and high costs. Traditional scoring methods to help clinicians deciding whether a patient is ready for discharge have failed to meet expectations, paving the way for machine learning based approaches. Freely available datasets such as MIMIC-III have served as benchmarking media to compare such tools. We used the OMOP-CDM version of MIMIC-III (MIMIC-OMOP) to train and evaluate a lightweight tree boosting method to predict readmission in ICU at different time points after discharge (3, 7, 15 and 30 days), outperforming existing solutions with an AUROC of 0.802 (SD=0.011) and a recall of 0.837 (SD=0.016) for 3-days readmission.

ICU readmission | MIMIC | OMOP-CDM | gradient boosting
Correspondence: ***dinh-phong.nguyen@aphp.fr***

Introduction

Recent studies have shown that readmission in the intensive care unit (ICU) is associated with poor clinical outcomes, increased length of ICU and hospital stay, and high costs (1, 2). One of the main reasons for ICU readmission that has been identified is premature discharge (3); in fact the transfer of patients from an ICU to a general hospital ward represents a high-risk event, and thus the decisions about which patients are ready to be discharged are daily struggles for ICU clinicians (4). Other studies have shown that determining the best timing for ICU discharge is usually based on subjective intuitions and that readmission prediction tools can help physicians in this endeavor, provided their performance and ease of adoption (5, 6). As traditional scores based on logistic regression or Cox proportional hazards models such as the Stability and Workload Index for Transfer score (SWIFT) or the LACE index have failed to meet expectations (6–10), numerous prediction models using machine learning have been proposed in the recent past, several of which trained and evaluated on the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC-II or MIMIC-III) open database (11–15).

MIMIC-III is a large ICU EHR database widely accessible to researchers internationally under a data use agreement, allowing clinical studies to be reproduced and benchmarked (16, 17). In order to make multicenter ressearch possible, a valuable effort has been made to convert MIMIC-III to the Observational Medical Outcomes Partnership common data model (OMOP-CDM), which provides structural and conceptual models relying on international reference terminologies (18, 19). For the sake of reproducibility and ease of subsequent implementation in other centers using the same data model, we chose to use the OMOP-CDM version of MIMIC-III (MIMIC-OMOP), for which documentation and a mapping Extract-Transform-Load (ETL) process are freely available.

Related works

Among the numerous works aiming to provide decision-making tools for ICU clinicians at discharge time, two in particular caught our attention in terms of performance and similarity to our own.

Lin et al. (12) proposed an advanced neural network for 30-day ICU readmission prediction (LSTM-CNN based model) achieving an Area Under Curve of the Receiver Operating Characteristic (AUROC) metric of 0.791 on MIMIC-III, using chart events 48h time series, diagnostic ICD-9 codes embeddings, and demographic information of the patients. The authors claim to offer higher accuracy and sensitivity compared to existing solutions.

Pakbin et al. (13) trained a simpler and more interpretable gradient boosting model (XGBoost) for predicting risk of ICU bounceback and readmission at a variety of time points using MIMIC-III, achieving AUROC of 0.76 and 0.75 for 72h and 30-days ICU readmission respectively with chart events time series, ICD-9 codes indicators, as well as admission, demographic and length-of-stay information of the patients.

Methods

Data and patients. MIMIC-III integrates deidentified, comprehensive clinical data of patients admitted from 2001 to 2012 at the Beth Israel Deaconess Medical Center in Boston, Massachusetts. We restricted our analyses to ICU stays of patients over 18 years old, ending up with a dataset of 55136 stays. Variables used in the model were provenance from a surgery ward, age, gender, length of stay, current count of ICU visits for the same patient, and three values for a number of measures and blood tests: the first entry for a given stay, the last one (which could be the same as the first) and the absolute difference between the two. Those measures were total glasgow coma scale (GCS), motor GCS, verbal GCS, eye movement GCS, systolic blood pressure, heart rate, respiratory rate, body temperature, oxygen saturation, oxygen inspired fraction, body weight, urine output, serum bicarbonate, serum urea, total bilirubin, serum sodium, serum potassium, serum creatinine, blood platelets, hemoglobinemia, blood hematocrit, blood leukocytes, serum lactates, blood PH, blood glucose and the International Normalized Ratio (INR). Three variables were extracted from

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2019.12.26.19015909; this version posted December 30, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

medical and nurses text notes: history of AIDS, metastatic cancer and/or of advanced hematologic condition (myeloma, lymphoma or leukemia). Finally, one feature counting the number of available values for all the previous variables was added. Missing data were then imputed via multivariate imputation (20), with a Bayesian ridge regression as the estimator (21).

We used a similar outcome definition to by Lin et al. (12), where positive cases were regarded as the ones where the patients could benefit from a prediction of readmission before being transferred or discharged: patients who were transferred or discharged but returned to ICU or died before a defined time limit (3, 7, 15 or 30 days).

Models. Several models were tested in a screening phase. Among those, XGBoost, a gradient tree boosting method that is widely used by data scientists to achieve state-of-the-art results on many machine learning challenges, was consistently more performant than the others and was thus selected as the prediction model for this study (22). Gradient boosting methods work by iteratively fitting "weak" models to the residuals of the previous model, and adding the newly estimated residuals to the previous model's prediction, thus forming a "new" prediction, and so on until a stopping criterion is met. XGBoost implements this algorithm with decision trees, an additional custom regularization term in the objective function, and several tweaks to optimize speed and performance.

Evaluation. Combining both clinical usefulness and comparability with other studies, area under the receiver operating characteristic curve (AUROC) was chosen as the main evaluation metric. Precision (=positive predicted value), recall (=sensitivity) and F1-score, the harmonic mean between the last two, were also reported. The probability threshold to classify a subject was chosen to be the class imbalance rate for each evaluated outcome, equal to the proportion of positive cases among the population. Stratified cross-validation was used in two ways; 5-fold for hyperparameter selection in a grid-search setting, and 10-fold for model evaluation.

Calibration of the models were assessed by separating the predicted probabilities over the whole dataset into deciles, and assessing the proportion of real outcomes in each bin. A model is said to be well calibrated when each bin's true outcomes proportion is close to the bin's predicted probabilities, resulting in a calibration curve close to the diagonal line when plotted.

Contributions of each feature to the final model were reported via feature importance; the more a feature is used to separate groups within the model's trees, the higher its relative importance. It is important to note that this might not necessarily mean that features with a higher importance are significantly associated with the outcome in real life, but it is nonetheless a good way to assess what the model has learned.

Reproducibility. All code used to produce this work, essentially written in Python, is available at http://github.com/deepphong/icu-readmission, accompanied by a step by step example Jupyter notebook and generic functions to exploit any database in OMOP format for other use cases.

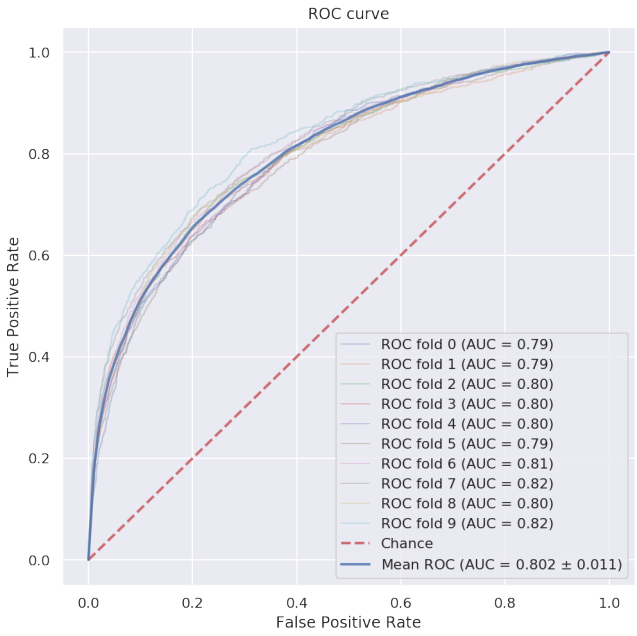

Fig. 1. Receiver Operating Characteristic (ROC) response for 3-days readmission prediction of different datasets, created from 10-fold cross-validation, showing how the output is affected by changes in the training data, and how different the splits are from one another and from the mean ROC curve and area under curve (AUC).

Results

The stratified 10-fold cross-validation results are reported in Table1 for each metric and outcome. Overall, our model consistently outperforms the results previously reported by Lin et al. (12) for 30-days readmission (mean AUROC 0.795 vs. 0.791, mean recall 0.778 vs. 0.742) and Pakbin et al. (13) for 3-days, 7-days and 30-days readmission (mean AUROC 0.80 vs. 0.76, 0.81 vs. 0.77 and 0.80 vs. 0.75 respectively; mean F1-score 0.36 vs. 0.22, 0.44 vs. 0.32 and 0.50 vs. 0.37 respectively). Performance was stable across all ten folds, as seen in Figure 1 for 3-days readmission prediction (mean AUROC = 0.802, range across folds = [0.79; 0.82]). Calibration of the model was overall very good, as visually assessed in Figure 2, with a near perfect fit towards the extremes, meaning that the model is more frequently right the more confident it is in making its predictions. Figure 3 shows the model's feature importance for 3-days readmission prediction. Apart from the expected high reliance on elements of the Glasgow Coma Scale (GCS), the most discriminative feature by far seems to be the one representing the number of non-missing features of an individual for a given stay; this feature is akin to the number of different measurements had by an individual for a given stay, and is plausibly correlated with the severity of the patient's condition.

Receiver Operating Characteristic (ROC) curves and calibration plots for all evaluated ICU readmission times are available in Appendix.

medRxiv preprint doi: https://doi.org/10.1101/2019.12.26.19015909; this version posted December 30, 2019. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

	3-days readmission (CIR=15.8%)	7-days readmission (CIR=19.7%)	15-days readmission (CIR=23.5%)	30-days readmission (CIR=27.3%)
Mean AUROC (SD)	0.802 (0.011)	0.809 (0.008)	0.803 (0.007)	0.795 (0.004)
Mean recall (SD)	0.837 (0.016)	0.821 (0.015)	0.799 (0.011)	0.778 (0.013)
Mean precision (SD)	0.421 (0.01)	0.494 (0.012)	0.547 (0.013)	0.587 (0.012)
Mean f1 (SD)	0.363 (0.024)	0.439 (0.022)	0.482 (0.015)	0.502 (0.009)

Table 1. Mean and standard deviation (SD) of the metrics measuring model performance for different outcomes, computed over a stratified 10-fold cross-validation. CIR = class imbalance ratio (proportion of positive cases among the population).

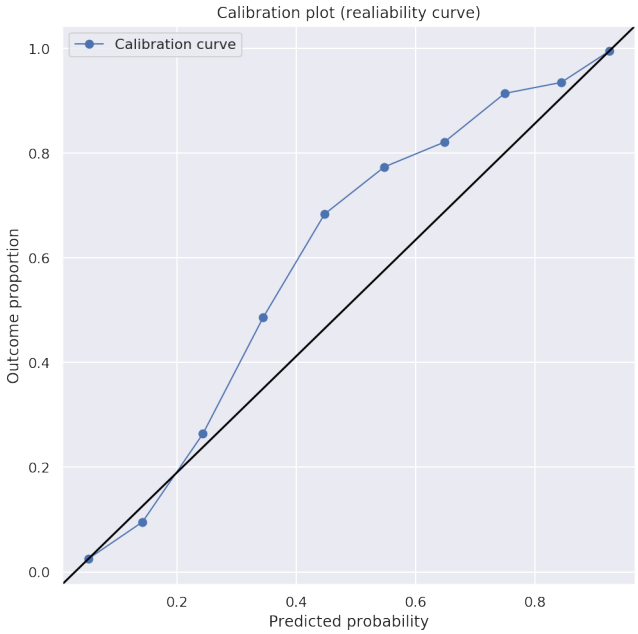

Fig. 2. Calibration plot showing the actual proportion of readmission in each decile of predicted probability of readmission.

Discussion

While this work improves on existing solutions for ICU readmission prediction as measured by commonly used metrics, several points still need to be addressed.

The prediction model was trained and evaluated on MIMIC-OMOP, and as of this time, no external validation has been conducted yet. To facilitate this process, all code needed to reproduce the results has been made freely available, and efforts have been put into its ease-of-transfer on any other electronic health records (EHR) database using the OMOP-CDM standard. External validation is planned to be conducted on several French EHR databases.

Further work needs to be done to integrate and evaluate such prediction models in real life settings. The choice of developing on common health data standards such as OMOP-CDM and/or HL7's Fast Health Interoperability Resources (FHIR) (23) is a step forward in this direction.

Apart from the performance, an interesting finding was the importance of including a variable accounting for the available measures among the ones selected for the model. This shows that the missing measures were missing not at random (MNAR), and the rationale behind this is that patients with a poorer prognosis usually have more tests and measurements done to them. This feature could possibly be indirectly correlated with the care providers' overall feeling of the patient's current state, and we postulate that existing models could potentially be improved by adding similar information.

Overall, when compared with previously published alternatives, our solution appears to be more accurate and sensitive, but also simpler, easier-to-implement and lighter than others such as deep learning based models and/or time-series based prediction methods.

Conclusion

In this study, we proposed a model based on a tree boosting method to predict ICU readmission at 3, 7, 15 and 30 days using data of patients at discharge on the freely available MIMIC-III database. With a 3-days readmission AUROC of 0.802 (SD=0.011) and recall of 0.837 (SD=0.016), or solution outperforms existing ones and has the advantage of having been conceived with the OMOP-CDM standard, allowing for easier external validation and implementation.

ACKNOWLEDGEMENTS

This work has been made possible thanks to the support of the WIND-DSI department of AP-HP hospitals and the DIM department of Bicetre hospital, especially Christel DANIEL and Marie FRANK. Special thanks to Julien DUBIEL from WIND-DSI for his constant help with logistics.

Bibliography

1. Sydney Brown, Sarah Ratcliffe, Jeremy Kahn, and Scott Halpern. The Epidemiology of Intensive Care Unit Readmissions in the United States. *American Journal of Respiratory and Critical Care Medicine*, 185(9):955–964, May 2012. ISSN 1073-449X. doi: 10.1164/rccm.201109-1720OC.
2. Carolina R. Ponzoni, Thiago D. Corrêa, Roberto R. Filho, Ary Serpa Neto, Murillo S. C. Assunção, Andreia Pardini, and Guilherme P. P. Schettino. Readmission to the Intensive Care Unit: Incidence, Risk Factors, Resource Use, and Outcomes. A Retrospective Cohort Study. *Annals of the American Thoracic Society*, 14(8):1312–1319, May 2017. ISSN 2329-6933. doi: 10.1513/AnnalsATS.201611-851OC.
3. Uchenna R. Ofoma, Yue Dong, Ognjen Gajic, and Brian W. Pickering. A qualitative exploration of the discharge process and factors predisposing to readmissions to the intensive care unit. *BMC Health Services Research*, 18(1):6, January 2018. ISSN 1472-6963. doi: 10.1186/s12913-017-2821-z.
4. Daniel Niven, Jaime Bastos, and Henry Stelfox. Critical Care Transition Programs and the Risk of Readmission or Death After Discharge From an ICU: A Systematic Review and Meta-Analysis*. *Critical Care Medicine*, 42(1):179–187, January 2014. ISSN 0090-3493. doi: 10.1097/CCM.0b013e3182a272c0.
5. Claudia-Paula Heidegger, Miriam M. Treggiari, Jacques-André Romand, and and the Swiss ICU Network. A nationwide survey of intensive care unit discharge practices. *Intensive Care Medicine*, 31(12):1676–1682, Dec 2005. ISSN 1432-1238. doi: 10.1007/s00134-005-2831-x.
6. Uchenna R. Ofoma, Subhash Chandra, Rahul Kashyap, Vitaly Herasevich, Adil Ahmed, Ognjen Gajic, Brian W. Pickering, and Christopher J. Farmer. Findings from the implementation of a validated readmission predictive tool in the discharge workflow of a medical intensive care unit. *Annals of the American Thoracic Society*, 11(5):737–743, June 2014. ISSN 2325-6621. doi: 10.1513/AnnalsATS.201312-436OC.
7. Marc Kastrup, Robert Powollik, Felix Balzer, Susanne Röber, Robert Ahlborn, Vera von Dossow-Hanfstingl, Klaus D. Wernecke, and Claudia D. Spies. Predictive ability of the stability and workload index for transfer score to predict unplanned readmissions after ICU discharge. *Critical Care Medicine*, 41(7):1608–1615, July 2013. ISSN 1530-0293. doi: 10.1097/CCM.0b013e31828a217b.

medRxiv preprint doi: https://doi.org/10.1101/2019.12.26.19015909; this version posted December 30, 2019. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

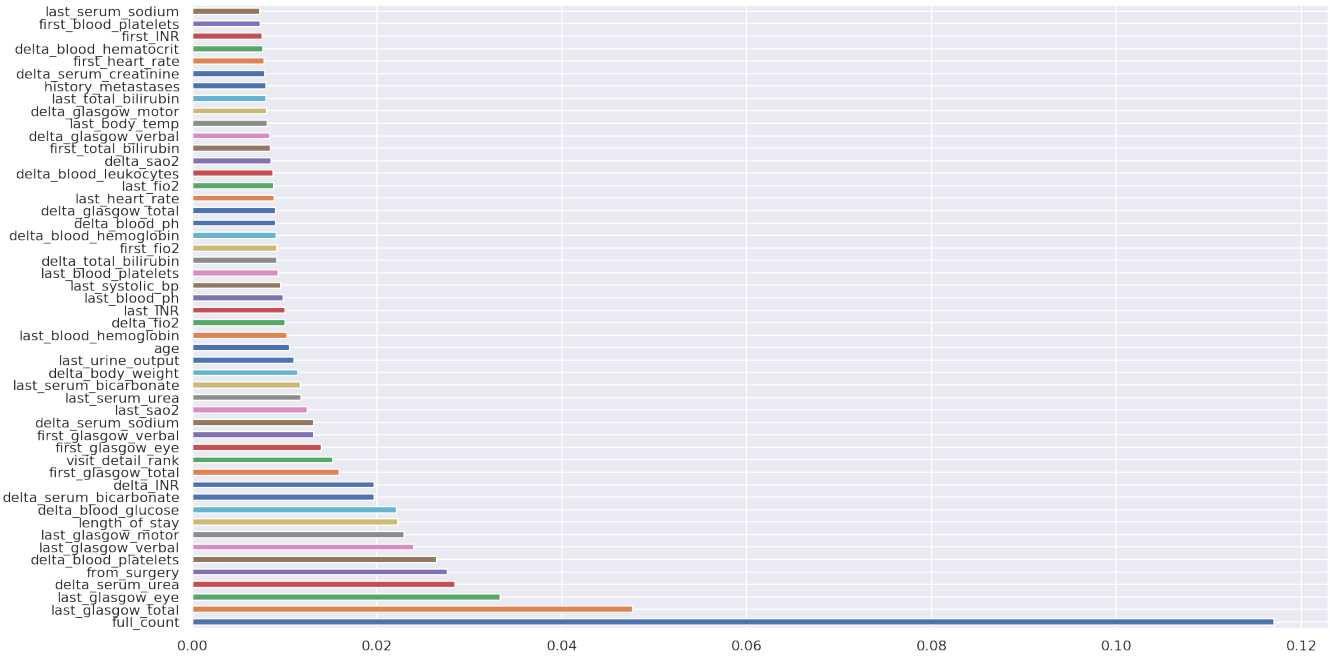

Fig. 3. Feature importance of the model for 3-days readmission prediction. Each feature prefixed by "first_", "last_" and "delta_" represent the first measure in the current stay, the last one and the difference between the two respectively. "full_count" represents the amount of non-missing features of an individual for a given stay.

8. Regis Goulart Rosa, Cintia Roehrig, Roselaine Pinheiro de Oliveira, Juçara Gasparetto Maccari, Ana Carolina Peçanha Antônio, Priscylla de Souza Castro, Felippe Leopoldo Dexheimer Neto, Patrícia de Campos Balzano, and Cassiano Teixeira. Comparison of Unplanned Intensive Care Unit Readmission Scores: A Prospective Cohort Study. *PloS One*, 10(11):e0143127, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0143127.
9. Devan Kansagara, Honora Englander, Amanda Salanitro, David Kagen, Cecelia Theobald, Michele Freeman, and Sunil Kripalani. Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA*, 306(15):1688–1698, October 2011. ISSN 0098-7484. doi: 10.1001/jama.2011.1515.
10. Carl van Walraven, Irfan A. Dhalla, Chaim Bell, Edward Etchells, Ian G. Stiell, Kelly Zarnke, Peter C. Austin, and Alan J. Forster. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *CMAJ : Canadian Medical Association Journal*, 182(6):551–557, April 2010. ISSN 0820-3946. doi: 10.1503/cmaj.091117.
11. Ye Xue, Diego Klabjan, and Yuan Luo. Predicting ICU readmission using grouped physiological and medication trends. *Artificial Intelligence in Medicine*, 95:27–37, April 2019. ISSN 1873-2860. doi: 10.1016/j.artmed.2018.08.004.
12. Yu-Wei Lin, Yuqian Zhou, Faraz Faghri, Michael J. Shaw, and Roy H. Campbell. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLOS ONE*, 14(7):e0218942, July 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0218942.
13. A. Pakbin, P. Rafi, N. Hurley, W. Schulz, M. Harlan Krumholz, and J. Bobak Mortazavi. Prediction of ICU Readmissions Using Data at Patient Discharge. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4932–4935, July 2018. doi: 10.1109/EMBC.2018.8513181.
14. J. Venugopalan, N. Chanani, K. Maher, and M. D. Wang. Combination of static and temporal data analysis to predict mortality and readmission in the intensive care. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2570–2573, July 2017. doi: 10.1109/EMBC.2017.8037382.
15. A. S. Fialho, F. Cismondi, S. M. Vieira, S. R. Reti, J. M. C. Sousa, and S. N. Finkelstein. Data mining using clinical physiology at discharge to predict ICU readmissions. *Expert Systems with Applications*, 39(18):13158–13165, December 2012. ISSN 0957-4174. doi: 10.1016/j.eswa.2012.05.086.
16. Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3, May 2016. ISSN 2052-4463. doi: 10.1038/sdata.2016.35.
17. Hrayr Harutyunyan, Hrant Khachatrian, David C. Kale, Greg Ver Steeg, and Aram Galstyan. Multitask learning and benchmarking with clinical time series data. *Scientific Data*, 6(1):96, 2019. ISSN 2052-4463. doi: 10.1038/s41597-019-0103-9.
18. George Hripcsak, Jon D. Duke, Nigam H. Shah, Christian G. Reich, Vojtech Huser, Martijn J. Schuemie, Marc A. Suchard, Rae Woong Park, Ian Chi Kei Wong, Peter R. Rijnbeek, Johan van der Lei, Nicole Pratt, G. Niklas Norén, Yu-Chuan Li, Paul E. Stang, David Madigan, and Patrick B. Ryan. Observational health data sciences and informatics (ohdsi): Opportunities for observational researchers. *Studies in health technology and informatics*, 216:574–578, 2015. ISSN 1879-8365.
19. J. Marc Overhage, Patrick B. Ryan, Christian G. Reich, Abraham G. Hartzema, and Paul E. Stang. Validation of a common data model for active safety surveillance research. *Journal of the American Medical Informatics Association : JAMIA*, 19(1):54–60, 2012. ISSN 1527-974X. doi: 10.1136/amiajnl-2011-000376.
20. Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software, Articles*, 45(3):1–67, 2011. ISSN 1548-7660. doi: 10.18637/jss.v045.i03.
21. J.L. Schafer. *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 1997. ISBN 9781439821862.
22. Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.
23. D. Bender and K. Sartipi. Hl7 fhir: An agile and restful approach to healthcare information exchange. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 326–331, June 2013. doi: 10.1109/CBMS.2013.6627810.