Postdoctoral position, Loria (CNRS, Inria, Université de Lorraine): "Quantifying and predicting the heterogeneous treatment effect of pharmacogenomic drugs"

Adrien Coulet, Anne Gégout-Petit

November 2019

1 Introduction

Precision medicine aims at improving clinical care using available information on individuals such as genetics, lifestyle, and environment [7]. Considering such information may help in prescribing the right drug at the right dose, and in the process reduce adverse drug reactions (ADRs), which are estimated to account for one-third of hospital adverse events and approximately 280,000 hospital admissions annually in the US [3, 11]. The inter-individual **variability in drug responses**, including ADRs, may have diverse causes such as: patient conditions, *e.g.*, a renal dysfunction impacts response to renally excreted drugs; drug interactions, *e.g.*, fluoxetin inhibits the effect of tamoxifen; drug-food interactions *e.g.*, grapefruit inhibits drug metabolism enzymes, causing drug toxicity; genetics *e.g.*, a genomic variation in CYP3A4 enzyme impacts drug responses. These few examples are well known because their impact on drug response is strong enough, however many factors are suspected to have smaller impact, which are combined and are hard to isolate one from the other. The variety of factors, both known, suspected or unknown, makes it challenging for health institutions to take proper precautions [10].

Electronic Health Records (EHRs) offer unprecedented opportunities for using patient data to study variable patient outcomes, including drug response [8, 9]. A new trend is to use the EHR recorded phenotypes as surrogate markers of individual profiles that may be associated to differential clinical outcomes. An advantage of available phenotypes in EHRs is to provide an "integrated view" of both genetic and environmental factors impacting patients. Along with Stanford University and the CHRU Nancy, we obtained initial results while using phenotypic data of EHR to predict the variable response to drug exposure [4]. We succeeded in demonstrating the feasibility of using historical data of an individual, recorded in EHRs prior to the drug exposure, to predict a reduced drug-dosing event.

2 Postdoctoral research subject

The aim of this postdococtoral project is to build upon previous results by first developing approaches that from EHR data identify without *a priori* groups of patients with distinct response profiles to particular drugs. Second, discovered groups will be used to identify predictors of drug response profiles.

We want to study the use of the causal inference framework [6, 5], and in particular of double robust approaches to identify groups with heterogeneous treatment effect, or in other words with significantly different drug response profiles. Through others, Athey and Imbens [1] introduced Causal Trees for subgroup analyses. Those are regression trees with modified splitting rules that maximise the difference in treatment effect between groups. Causal Trees have been reused to propose an ensemble method named Causal Forest [12, 2] that has the advantage of being non-parametric and consistent in many settings. These models will allow us to estimate the causal effect of the phenotype on the drug response.

A second objective of the project is to develop high performance predictive models for drug response profiles that we aim at identifying with causal methods. These models can be seen as classifiers that assign individuals to a specific profile. A first challenge here is to develop real predictive models, *i.e.*, models that are trained only on data collected prior to the prescription of the drug. A second challenge is to identify a subset of good predictive features that may help in interpreting group belonging and heterogeneous drug responses.

The set of drug studies will first be pharmacogenomic drugs, *i.e,* drugs known to present a variability in the population for genomic reasons, and will potentially be extended in a second time to other drugs of interest.

3 Context of the postdoctoral position

- The postdoc will be at the Loria in **Nancy, France**
- is funded for **1 year**, jointly by the ANR project PractiKPharma, and the I-site LUE (Lorraine University of Excellence)
- start date is flexible, but preferably not later than March 1st, 2020
- is **extendable** for a second year.

The postdoc will work with Adrien Coulet (Orpailleur team, Loria, Inria, Nancy) and Anne Gégout-Petit (Bigs team, Institut Elie Cartan, Inria, Nancy)

The project is also an opportunity to maintain an existing collaboration with the Shah Lab at Stanford University. Experiments will be conducted first with data from the Shah Lab. Accordingly few travels to Stanford will be planed.

4 Contacts

Adrien Coulet, adrien.coulet@loria.fr, +33 3 54 95 86 38
Anne Gégout-Petit, anne.gegout-petit@univ-lorraine.fr,+33 3 72 74 53 78

References

[1] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *PNAS*, 113(27):7353–7360, 2016.

[2] Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *Annals of Statistics*, 47(2), 2019.

[3] Florence T. Bourgeois, Michael W. Shannon, Clarissa Valim, and Kenneth D. Mandl. Adverse drug events in the outpatient setting: an 11-year national analysis. *Pharmacoepid. and Drug Safety*, 19(9), 2010.

[4] Adrien Coulet, Nigam H. Shah, Maxime Wack, Mohammad Chawki, Nicolas Jay, and Michel Dumontier. Predicting the need for a reduced drug dose, at first prescription. *Scientific Reports*, 8(1), October 2018.

[5] Miguel A. Hernán and James M. Robins. *Causal Inference*. Chapman & Hall/CRC, Boca Raton, FL, USA, 2019.

[6] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York, NY, USA, 2015.

[7] J. Larry Jameson and Dan L. Longo. Precision medicine – personalized, problematic, and promising. *New England Journal of Medicine*, 372(23):2229–2234, 2015. PMID: 26014593.

[8] Peter B. Jensen, Lars J. Jensen, and Soren Brunak. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet.*, 13(6):395–405, 2012.

[9] A Neuraz, L Chouchana, et al. Phenome-wide association studies on a quantitative trait: Application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics. *PLoS Comp Bio.*, 9(12), 2013.

[10] Walter F. Stewart, Nirav R. Shah, Mark J. Selna, Ronald A. Paulus, and James M. Walker. Bridging the inferential gap: The electronic health record and clinical evidence. *Health Affairs*, 26(2):w181–w191, 2007.

[11] U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. National action plan for adverse drug event prevention., 2014.

[12] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 2018.