Equipe Inria-Inserm
PreMeDICAL: Precision Medicine by Data Integration and Causal Learning

Inria Sophia Antipolis - Méditerranée, Antenne de Montpellier.

Institut Desbrest d’Épidémiologie et de Santé Publique (IDESP): UMR UA11 Inserm - Université de Montpellier (UM). Focus on epidemiology of chronic non communicable diseases.
Composition of the team

- **Julie Josse (PI) 100%**: Inria. Topics: Missing values, Dimensionality reduction, matrix completion, causal inference, R statistical software
- **Pascal Demoly 50 %**: PU-PH, Director of Idesp. Respiratory physician, Professor of pulmonology University Hospital, head of the pulmonology dept
- **Pierre Lafaye de Micheaux 50 %**: Assistant professor, UPVM3. Topics: Measures of dependences, exploratory analyses, R statistical software
- **Nicolas Molinari 40 %**: PU-PH, Co-director of Idesp. Professor in biostatistics at the University Hospital, head of the statistics department

5 PhD students Marie Beclin (NM, PLM), Pan Zhao (JJ), Maxime Fosset (NM, JJ), Margaux Zaffran (EDF), Charlotte Voinot (Sanofi). Postdoc: Jeff Naf

⇒ *Interdisciplinary* team with clinical, bio-stat & machine learning (ML) skills
**PreMedICaL research axes**

**Personalized medicine by optimal prescription of treatment**

- causal inference techniques for (dynamic) policy learning
  ⇒ who to treat and when
- leverage both randomized control trials (RCTs) and observational data
  ⇒ launch a drug without running RCTs
  ⇒ rethink evidence needed to bring treatments to the market faster

**Personalized medicine by integration of different data sources**

- relevance of each data source from different scales, measures of dependencies - distributed computation
- solutions to handle missing values: complex structure of missing values, prediction with uncertainty quantification

⇒ Push methodological innovation up to patients, clinicians, regulators
⇒ Collaborative effort: leveraging ML, data, clinical expertise and existing recommendations
Missing values in multi-source, multi-scale data

<table>
<thead>
<tr>
<th></th>
<th>Clinical Data</th>
<th>Biological Data</th>
<th>Questionnaire on lifestyle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital 1</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>n_1</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Hospital 2</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>n_2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital K</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>n_K</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Classical methodologies are not designed to handle high-dimensional data with selection bias and informative missing data.
Short - mid terms objectives

From a methodological point of view

- innovative methods to handle missing values (multiple imputation, low-rank approximation + optimal transport, etc.)
- new development in causal inference (survival causal inference, sensitivity analyses)
- provide easy-to-use tools (such as R package) and reproducible pipelines to allow for direct deployment by stakeholders

From a patient/medical point of view

- personalized benefit of treatment (over time)
- identify subgroup of patients
- adoption by the medical community of advanced techniques (end users: biostatisticians in DIM)

PreMeDICAL: bio-statistics, ML, methodological specificities, rapid transfer through software. First application to allergy, then other diseases
Long terms objectives

From a methodological point of view

- Design future clinical trials (PEPR). Software as decisions tools
- New research axis: Federated learning - Aurélien Bellet. (+CR)
- **Exposome** - ExposUM: European project Horizon: links between early respiratory infections and paediatric asthma onset to build predictive, preventive & curative tools (microbiome, pollution, diet, epigenetic, clinical, biological, sociological data, etc.)

From a patient/medical point of view

- give patient better care and early access to innovation
- better management of resources
- guide decisions made by investigators, sponsors and authorities

Relationship with companies, pharmaceutical (ALK, Astrazeneca, etc). International network (Stanford, ML for health) on causality, communauty on generalization (software), scientific mediation, etc.