Contents

Project-Team SISTM 1

1 Team members, visitors, external collaborators 2

2 Overall objectives 3

3 Research program 4

3.1 Axis 1 - High-Dimensional Statistical Learning 4
3.2 Axis 2 - Mechanistic learning 5
3.3 Axis 3 - Translational vaccinology 6

4 Application domains 6

4.1 Systems Biology and Translational medicine 6
4.2 HIV immunotherapies 7
4.3 Translational vaccinology 7

5 Highlights of the year 8

5.1 Ebola vaccine development 8
5.2 Response to COVID19 8
5.3 Awards 8

6 New software and platforms 8

6.1 New software 8
6.1.1 PheVIS 8
6.1.2 sGBJ 9
6.1.3 FrechForest 9
6.1.4 EpidemioOptim 9
6.1.5 VASI_Cyto 9
6.1.6 ccdf 10

7 New results 10

7.1 Methods for transcriptomic studies 10
7.2 Translational vaccinology 10
7.2.1 Vaccine clinical trials 10
7.2.2 Covid treatment trials 10
7.2.3 Systems immunology analyses 10
7.2.4 Methodological developments for vaccine trials 11

8 Partnerships and cooperations 11

8.1 International initiatives 11
8.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program 11
8.1.2 Participation in other International Programs 12
8.2 International research visitors 12
8.2.1 Visits of international scientists 12
8.3 European initiatives 12
8.3.1 FP7 and H2020 projects 12
8.3.2 Other european programs/initiatives 12
8.4 National initiatives 13
8.4.1 Expert Appraisals 14
8.4.2 Various Partnership 14
9 Dissemination

9.1 Promoting scientific activities ................................. 15
  9.1.1 Scientific events: organisation ............................ 15
  9.1.2 Scientific events: selection ............................... 15
  9.1.3 Journal .................................................... 15
  9.1.4 Invited talks .............................................. 16
  9.1.5 Scientific expertise ....................................... 16
  9.1.6 Research administration .................................. 17

9.2 Teaching - Supervision - Juries ............................... 17
  9.2.1 Teaching .................................................. 17
  9.2.2 Supervision ............................................... 18
  9.2.3 Juries ..................................................... 19

9.3 Popularization .................................................. 20
  9.3.1 Articles and contents .................................... 20
  9.3.2 Education ................................................ 20
  9.3.3 Interventions ............................................. 20

10 Scientific production ............................................. 21
  10.1 Major publications .......................................... 21
  10.2 Publications of the year .................................... 22
  10.3 Other ....................................................... 27
  10.4 Cited publications .......................................... 27
Project-Team SISTM

Creation of the Project-Team: 2015 January 01

Keywords

Computer sciences and digital sciences
A3.1.10. – Heterogeneous data
A3.3.2. – Data mining
A3.3.3. – Big data analysis
A3.4.1. – Supervised learning
A3.4.2. – Unsupervised learning
A3.4.4. – Optimization and learning
A3.4.5. – Bayesian methods
A6.1.1. – Continuous Modeling (PDE, ODE)
A6.2.4. – Statistical methods
A6.3.1. – Inverse problems
A6.3.4. – Model reduction
A6.4.2. – Stochastic control
A9.2. – Machine learning

Other research topics and application domains
B1.1. – Biology
B1.1.5. – Immunology
B1.1.7. – Bioinformatics
B1.1.10. – Systems and synthetic biology
B2.2.4. – Infectious diseases, Virology
B2.2.5. – Immune system diseases
B2.3. – Epidemiology
B2.4.1. – Pharmaco kinetics and dynamics
B2.4.2. – Drug resistance
B9.5.6. – Data science
B9.8. – Reproducibility
1 Team members, visitors, external collaborators

Research Scientists
- Boris Hejblum [INSERM]
- Melanie Prague [Inria]

Faculty Members
- Rodolphe Thiébaut [Team leader, Univ de Bordeaux and Univ Hospital, Professor, HDR]
- Marta Avalos Fernandez [Univ de Bordeaux, Associate Professor, HDR]
- Robin Genuer [Univ de Bordeaux, Associate Professor, HDR]
- Edouard Lhomme [Univ de Bordeaux and Univ Hospital, Associate Professor since Sept 2021]
- Laura Richert [Univ de Bordeaux and Univ Hospital, Professor, HDR]
- Linda Wittkop [Univ de Bordeaux and Univ Hospital, Associate Professor, HDR]

Post-Doctoral Fellows
- Louis Capitaine [Univ de Bordeaux, Until Dec 2021]
- Quentin Clairon [INSERM]
- Laura Villain [INRIA, Until Sep 2021]

PhD Students
- Marie Alexandre [Inria]
- Houreratou Barry [INSERM]
- Thomas Ferte [Bordeaux Univ Hospital, From Nov 2021]
- Iris Beatrice Ganser [INSERM]
- Marine Gauthier [Univ de Bordeaux, until Dec 2021]
- Benjamin Hivert [INSERM]
- Helene Savel [Ipsen, CIFRE]

Technical Staff
- Kalidou Ba [INSERM, Engineer, From Oct 2021]
- Melany Durand [INSERM, Engineer]
- Melanie Huchon [INSERM, Engineer, from Oct 2021]
- Laurent Lehot [INSERM, Engineer, until Jun 2021]
- Clement Nerestan [Inria, Engineer]
- Ndoh Penn [INSERM, Engineer, until Feb 2021]
- Maria Prieto Gonzalez [INSERM, Engineer, until Aug 2021]
- Myrtille Richard [INSERM, Engineer, until Jun 2021]
- Panthea Tzourio [INSERM, Engineer]
Interns and Apprentices

• Emma Avisou [INSERM, from Apr 2021 until Jun 2021]
• Marie Charlotte Ballion [INSERM, from Apr 2021 until Jun 2021]
• Clement Bonnet [INSERM, from Apr 2021 until Jun 2021]
• Marie Laure Charpignon [INSERM, from Jun 2021 until Sep 2021]
• Guillaume Coulaud [Univ Bordeaux, From May 2021 until Jul 2021]
• Celine El-Zein [INSERM, from Apr 2021 until Aug 2021]
• Mélanie Huchon [Inserm, From Mar 2021 until Aug 2021]
• Clemence Metayer [Inria, from Apr 2021 until Oct 2021]
• Abdelghani Neuhaus [Inria, from Feb 2021 until Aug 2021]
• Carole Vignals [INSERM, from Feb 2021 until Jun 2021]

Administrative Assistants

• Sandrine Darmigny [INSERM]
• Anton Ottavi [INSERM, Research funding coordinator]
• Audrey Plaza [Inria]

Visiting Scientists

• Gayo Diallo [Univ de Bordeaux, HDR]
• Jane Heffernan [INSERM, until Aug 2021]

External Collaborator

• Thomas Ferte [Bordeaux Univ Hospital, until Oct 2021]

2 Overall objectives

The two main objectives of the SISTM team are: i) to accelerate the development of vaccines by analyzing all the information available in early clinical trials and optimizing new trials ii) to develop new data science approaches to analyze and model big/omics data. The methods developed are relevant in many other applications than those encountered in the SISTM team. However, the focus devoted to vaccine development is justified by the importance of the objective from a public health point of view and a good knowledge of the application field that maximizes the relevance and the implementation of the methods developed.

This equilibrium between the methodological and the applied work reached over the last years is a fundamental motivation for each member of the SISTM team even though the background could be very different from one researcher to the other (e.g. math vs. public health). This equilibrium is maintained by the organization of the team as well as the collaborations established especially through the Vaccine Research Institute, Inserm and Inria. Hence, we are able to collaborate for a theoretical problem during the development of a new method (e.g. demonstration of the convergence of an estimator) as well as to translate the research outcomes (either new analytical methods or applied results) to clinicians and biologists first in our collaborative networks and then beyond.

The SISTM team benefits from a very rich ecosystem. Firstly, it is one of the rare team belonging to Inserm and Inria national institutes, which helps establishing collaboration as testified by the co-supervision of PhD Students and co-publications with other researchers belonging to Inserm teams or
Inria teams from the two research centres in Bordeaux. Secondly, the applications in clinical research are facilitated by the very close collaboration with Clinical Trial Units (CTUs): from the ANRS/VRI (CMG directed by Linda Wittkop), from Bordeaux Hospital (USMR directed by Laura Richert and previously by Rodolphe Thiébaut), from the international consortia linked to the Vaccine Research Institute (e.g. MRC in EHVA, Janssen in EBOVAC). Finally, the team is very much involved in teaching activities in the context of ISPED and the Graduate’s program Digital Public Health (directed by Rodolphe Thiébaut). Hence, two specialties of the Master of Public Health (Biostatistics and Public Health Data Science) are supervised by members of the SISTM team.

3 Research program

The team is organized in three research axes:

1. High Dimensional Statistical Learning (leader Boris Hejblum),
2. Mechanistic learning (leader Mélanie Prague),
3. Translational vaccinology (leader Laura Richert).

Our specific scientific objectives are:

3.1 Axis 1 - High-Dimensional Statistical Learning

- To uncover and visualize sparse signal and clustering structures from high-dimensional omics data through statistically sound methodologies.
- To infer cellular populations abundance from transcriptomics data with semi-supervised learning.
- To reduce data redundancy by i) high-dimensional reduction and ii) deconvolution methods.
- To develop and apply methods for discovering complex relationships between high-dimensional data (multiblock analysis for data integration).

The first step for each type of data is to infer the maximum of information from the generated data. In spite of being high-dimensional, data are generally not analyzed in full due to either their size and complexity. For example in cellular phenotyping data, only a set of a given combination of markers are used to measure the abundance of a pre-defined set of cell types. This approach precludes the discovery of new types of cells defined by new combinations of markers. This problem is even more prevalent with current high-throughput technologies that allows measurement of up to 100 markers on a single cell. We are thus developing statistical learning approaches for automated gating of cellular populations, for a comprehensive range of single-cell measurement technologies (e.g. Flow-Cytometry, CyTOF). We still need to strengthen annotation procedures and to liaise with research on ontologies.

However, measuring specific cells across many intra-cellular and surface markers requires a lot of blood, preferentially fresh. This makes it very difficult to implement such measurements to large sample sizes with many repeated measurements. This is why we want to study how feasible it would be to replace cell phenotyping by transcriptomics analysis in whole blood. This is a challenging exercise that go much beyond the initial works done on this topic using deconvolution approaches. We foresee two complementary ideas that could help us reaching this objective: improving the model for the deconvolution using more sophisticated statistical and machine learning (ML) models (as described below), and to improve the knowledge databases by using newly available single cell transcriptomics analyses. Single cell transcriptomics is a newly available method that provides a new type of information. However, its analysis requires new approaches to take into account several methodological difficulties such as heavily skewed distributions. We are already developing a new approach based on score test statistics that appears to provide much better control of type 1 and type 2 errors than currently available methods (dearseq, EdgeR). One of the methods that is particularly performant in our context of high dimensional data is the Random Forests. However, several limitations are present especially when dealing with repeated measures. Therefore, we are working on extension generalizing the available approaches of random forest for longitudinal data and we are also proposing new metrics (Frechet).

High-dimensional data such as omics data are often analyzed first by clustering observations to identify homogeneous sub-populations. It is then wildly popular to contrast the identified clusters to
highlight differentiating features, even though this re-use of the same data known as “double-dipping” is unsound according to statistical theory (leading to overfitting and optimism bias). We are currently developing on new approaches to this problem of post-clustering inference with a focus on (single-cell) RNA-seq transcriptomics data.

In each clinical trial, we are ending with various types of markers and the final objective is to describe the relationship between all markers. However, the high-dimensional setting makes the analysis much more complex. We have already developed extension of sparse Partial Least Squares and multiblock analyses but these solutions are not fully satisfying because they do not take into account missing data in X and Y properly, the behavior is often compromise in very high dimensional contexts that we are facing.

3.2 Axis 2 - Mechanistic learning

- To infer ordinary differential equations (ODE) systems parameters by using high dimensional data.

- To compare and implement control strategies through various approaches belonging to statistical control, stochastic control, reinforcement learning.

In this axis, we focus on Inference in population of mechanistic models. This modeling is constituted of three features: 1/ a mathematical model, which describes a phenomenon, 2/ a statistical model, which describes the variability that exists in data, and 3/ an observational model, which relates what is observable with the mathematical model. In each feature, new methodological development in this research program are described below:

1. A significant part of the work done in this axis consists in modeling dynamics of markers in immunology and vaccinology. This consists in reviewing the literature and discussing with clinicians to model at best the knowledge of a biological mechanism. For each new project, calibration and identifiability analysis are needed to check the feasibility of answering biological questions with mathematical tools (we have examples in the field of Ebola vaccine response, HIV antiretroviral interruption strategies, NIPAH infection, COVID19 epidemics in population...). An example of such theoretical work for the specific topic of humoral response to Ebola vaccine is available in Balelli et al. 2020.

2. When data are available over multiple subjects, using an ODE with mixed effects on parameters has the advantage to borrow information from most the informative subjects to estimate parameters of the other subjects. However, when maximum-likelihood based approaches are computationally intensive, thus it is very difficult to scale up when the complexity of the mathematical model gets higher. While we routinely use Stochastic Approximation Estimation Maximisation (SAEM) algorithm as implemented in Monolix software, we intend to develop new methods for estimation in these models following two tracks: 1/ using Kalman filtering approach (Colin et al. 2020, submitted) and 2/ using optimal control algorithms (Clairon et al. 2020, in preparation).

3. In vaccinology and immunology, various markers are measured. However, their observation is prone to error (either random or constrained by experiments) and to uncertainty due to the possibly misspecified understanding of a mechanism. Latent class models, such as in Proust-Lima et al. (2013), address how to build a latent process from observed variables to describe a phenomenon of interest. Other works such as Tadde et al. (2019), based on latent processes and Dynamic Bayesian Network aim at understanding a mechanism between multiple biomarkers and inferring a dynamical model. One limitation of these approach is that it mainly relies on correlations between covariates, but correlation is not causation. We aim at jointly estimating parameters in the dynamical system while estimating parameters of the observation model using latent class approaches and possibly lasso-type techniques.

All these methodological developments will help in answering open questions already raised by the data but will also pave the way to make use of high dimensional data in the observational model to account for information such as omics data. A particular example at hands is the use of repeated genes expression to track concentration of population of cells and use it as input in the mechanistic analysis. Once estimated with observed data, these models can be compared and used to implement control strategies. The comparison can be performed using model averaging, which refers to the practice of using several models for inferring parameters and for making predictions. It has been shown that model averaging leads to better performances than simple model selection (Gonçalves et al. 2019). Finally, it is possible to prospectively use the models to target optimal strategies of treatment or vaccination: we will continue building on Bayesian approaches such as Villain et al. 2019. However, given the complexity of
the epidemic dynamics (and the associated complexity of models), these pre-defined coarse strategies are bound to be sub-optimal, especially when considering that the problem is multi-objective and that strategies may be heterogeneous and multiscale (Halloran et al., 2008). We will investigate approaches which are more exhaustive and explore the space of solution such as one based on stochastic control (Pasin et al. 2019) or reinforcement learning which is indicated in high-dimensional non-stationary environments with uncertainty and partial observation of the state of the system (Mnih et al., 2015; Haarnoja et al., 2018).

3.3 Axis 3 - Translational vaccinology

- To accelerate vaccine development by in silico trials.
- To accelerate vaccine development by new adaptive designs.
- To accelerate vaccine development by in depth analysis of data generated in early clinical trials.

Beyond the analysis of clinical trials performed, especially in the context of the VRI, that bring new fundamental information for next trials and fully contribute to the development of vaccines (e.g. Sirima et al. Lancet Infectious diseases 2020, Pollard et al. Lancet Infectious diseases 2020, Jahnmatz et al. Lancet Infectious Diseases 2020, Barry et al. Plos Med. 2021 [19], Lévy et al. J Virol 2021 [33]), we want to capitalize with the research done in data science to accelerate vaccine development. We foresee three main ways. The first one is by the deep analysis of the data generated. One of the main hypotheses is that these analyses may help i) defining correlates of protection, that is surrogate markers reflecting the vaccine efficacy ii) stratifying the participants at baseline or very early to optimize the response to vaccine. A good example of the approach is our results with the Ebola rVSV vaccine where we found early markers measured at day 1 or 3 after the vaccine injection that were associated to the antibody response 6 months later (Rechtien et al. Cell report 2017).

The second project is to design next trials in a more efficient way. Adaptive designs are clearly relevant in the situation of vaccine development and we have already implemented such designs (VRI01, EHVA T01). However, we think that we can go much beyond and we want to explore new approaches such as using reinforcement learning. Although this is not our area of expertise, this will be done in collaboration with Inria teams (Flowers, Scool - E. Kaufmann-) and it is actually very much connected to the research performed for optimal control of treatment regimen developed in Axis 2. Last but not least, we want to explore in silico trials that is simulating potential trials to help selecting potential candidates for the next real trials. Opportunistically, we want to build on the mechanistic models developed in axis 2 to simulate new trials adding external information from already published studies. We already used this type of approach in the simple context of exogenous IL-7 therapy. Now, we want to use it to define best combination of vaccines platforms and adjuvants. Here we surmise on the fact that same vaccine vectors are used with many different antigens and combined with various adjuvants. For instance, the Adenovirus 26 vector is evaluated for HIV, Ebola and SARS-Cov2 vaccine.

In conclusion, the team is now well organized around three axes sharing a common objective. Longterm local, national and international collaborations have been settled with a part of fundings secured up to 2025. It is embarked in a double challenge of developing methods to deal with high dimensional data and a main application for accelerating vaccine development. The growing interest and expertise in machine learning and reinforcement learning approaches open new ways to reach these objectives.

4 Application domains

4.1 Systems Biology and Translational medicine

Biological and clinical researches have dramatically changed because of the technological advances, leading to the possibility of measuring much more biological quantities than previously. Clinical research studies can include now traditional measurements such as clinical status, but also thousands of cell populations, peptides, gene expressions for a given patient. This has facilitated the transfer of knowledge from basic to clinical science (from “bench side to bedside”) and vice versa, a process often called
"Translational medicine". However, the analysis of these large amounts of data needs specific methods, especially when one wants to have a global understanding of the information inherent to complex systems through an "integrative analysis". These systems like the immune system are complex because of many interactions within and between many levels (inside cells, between cells, in different tissues, in various species). This has led to a new field called "Systems biology" rapidly adapted to specific topics such as "Systems Immunology" [76], "Systems vaccinology" [75], "Systems medicine" [73]. From the data scientist point of view, two main challenges appear: i) to deal with the massive amount of data ii) to find relevant models capturing observed behaviors.

4.2 HIV immunotherapies

The management of HIV infected patients and the control of the epidemics have been revolutionized by the availability of highly active antiretroviral therapies. Patients treated by these combinations of antiretrovirals have most often undetectable viral loads with an immune reconstitution leading to a survival which is nearly the same to uninfected individuals [74]. Hence, it has been demonstrated that early start of antiretroviral treatments may be good for individual patients as well as for the control of the HIV epidemics (by reducing the transmission from infected people) [72]. However, the implementation of such strategy is difficult especially in developing countries. Some HIV infected individuals do not tolerate antiretroviral regimen or did not reconstitute their immune system. Therefore, vaccine and other immune interventions are required. Many vaccine candidates as well as other immune interventions (IL7, IL15) are currently evaluated. The challenges here are multiple because the effects of these interventions on the immune system are not fully understood, there are no good surrogate markers although the number of measured markers has exponentially increased. Hence, HIV clinical epidemiology has also entered in the era of Big Data because of the very deep evaluation at individual level leading to a huge amount of complex data, repeated over time, even in clinical trials that includes a small number of subjects.

4.3 Translational vaccinology

Vaccines are one of the most efficient tools to prevent and control infectious diseases, and there is a need to increase the number of safe and efficacious vaccines against various pathogens. However, clinical development of vaccines - and of any other investigational product - is a lengthy and costly process. Considering the public health benefits of vaccines, their development needs to be supported and accelerated. During early phase clinical vaccine development (phase I, II trials, translational trials), the number of possible candidate vaccine strategies against a given pathogen that needs to be down-selected in early clinical development is potentially very large. Moreover, during early clinical development there are most often no validated surrogate endpoints to predict the clinical efficacy of a vaccine strategy based on immunogenicity results that could be used as a consensus immunogenicity endpoint and down-selection criterion. This implies considerable uncertainty about the interpretation of immunogenicity results and about the potential value of a vaccine strategy as it transits through early clinical development. Given the complexity of the immune system and the many unknowns in the generation of a protective immune response, early vaccine clinical development nowadays thus takes advantage of high throughput (or "omics") methods allowing to simultaneously assess a large number of response markers at different levels ("multi-omics") of the immune system. Outside of the context of emergency vaccine development during a pandemic, this has induced a paradigm shift towards early-stage and translational vaccine clinical trials including fewer participants but with thousands of data points collected on every single individual. This is expected to contribute to acceleration of vaccine development thanks to a broader search for immunogenicity signals and a better understanding of the mechanisms induced by each vaccine strategy. However, this remains a difficult research field, both from the immunological as well as from the statistical perspective. Extracting meaningful information from these multi-omics data and transferring it towards an acceleration of vaccine development requires adequate statistical methods, state-of-the art immunological technologies and expertise, and thoughtful interpretation of the results. It thus constitutes research at the interface between disciplines: data science, immunology and vaccinology. Our main current areas of application here are early phase trials of HIV and Ebola vaccine strategies, in which we participate from the initial trial design to the final data analyses. We are also involved in the development of next-generation pan-Coronavirus vaccines.
5 Highlights of the year

5.1 Ebola vaccine development

Beyond the scientific publications including the results of EBL2002 trial (2 articles in Plos Med, 2021 [19] and Jan 2022), the Ad26.ZEBOV/MVA-BN-Filo developed with Janssen has been approved by FDA and EMA for emergency use on July 1st 2020. The submitted file included the model predictions of the duration of the vaccine response. In addition, WHO Prequalification for the use of the Ebola vaccine has been obtained on April 2021.

Another large randomized multi-arm Ebola vaccine trial (Prevac trial, evaluating three different Ebola vaccine strategies; Badio et al, Trials 2021 [16]) has had its primarily results available in 2021 and a manuscript has been submitted to the New England Journal of Medicine at the end of 2021.

5.2 Response to COVID19


5.3 Awards

For the whole work and coordination of the EBOVAC2 project, R Thiébaut has received the Prix Etoiles de l’Europe 2021, special mention Prix Etoiles de l’Europe 2021.

6 New software and platforms

List of new softwares.

6.1 New software

6.1.1 PheVIS

Name: PheVis: Automatic Phenotyping of Electronic Health Record at Visit Resolution

Keywords: Timeseries Prediction, Medical applications, Electronic Medical Records, Artificial intelligence

Functional Description: Using Electronic Health Record (EHR) is difficult because most of the time the true characteristic of the patient is not available. Instead we can retrieve the International Classification of Disease code related to the disease of interest or we can count the occurrence of the Unified Medical Language System. None of them is the true phenotype which needs chart review to identify. However chart review is time consuming and costly. ‘PheVis’ is an algorithm which is phenotyping (i.e identify a characteristic) at the visit level in an unsupervised fashion. It can be used for chronic or acute diseases. An example of how to use ‘PheVis’ is available in the vignette. Basically there are two functions that are to be used: ‘train_phevis()’ which trains the algorithm and ‘test_phevis()’ which get the predicted probabilities. The detailed method is described in preprint by Ferté et al. (2020) <doi:10.1101/2020.06.15.20131458>.

URL: https://CRAN.R-project.org/package=PheVis

Contact: Boris Hejblum

Partners: INSERM, CHU de Bordeaux
6.1.2 sGBJ

**Name:** Survival Extension of the Generalized Berk-Jones Test

**Keywords:** Biostatistics, Survival analysis, Cancer, Transcriptomics

**Functional Description:** Implements an extension of the Generalized Berk-Jones (GBJ) statistic for survival data, sGBJ. It computes the sGBJ statistic and its p-value for testing the association between a gene set and a time-to-event outcome with possible adjustment on additional covariates.

**URL:** [https://CRAN.R-project.org/package=sGBJ](https://CRAN.R-project.org/package=sGBJ)

**Contact:** Boris Hejblum

6.1.3 FrechForest

**Name:** Fréchet Random Forest

**Keyword:** Machine learning

**Functional Description:** The R package is used to train a Fréchet tree or random forest for a learning problem where the inputs can be a mixture of continuous or categorical variables, curves or images, and the output can be a continuous, categorical, curve or image variable.

**URL:** [https://github.com/Lcapitaine/FrechForest](https://github.com/Lcapitaine/FrechForest)

**Authors:** Louis Capitaine, Robin Genuer

**Contact:** Robin Genuer

6.1.4 EpidemiOptim

**Name:** EpidemiOptim: a toolbox for the optimization of control policies in epidemiological models

**Keywords:** Epidemiology, Optimization, Dynamical system, Reinforcement learning, Multi-objective optimisation

**Functional Description:** This toolbox proposes a modular set of tools to optimize intervention strategies in epidemiological models. The user can define or use a pre-coded epidemiological model to represent an epidemic. He/she can define a set of cost functions to define a particular optimization problem. Finally, given an optimization problem (epidemiological model and cost functions and action modalities), the user can define/reuse optimization algorithms to optimize intervention strategies that minimize the costs. Finally, the toolbox contains visualization and comparison tools. This allows to investigate various hypotheses easily.

**URL:** [https://github.com/flowersteam/EpidemiOptim](https://github.com/flowersteam/EpidemiOptim)

**Contact:** Cedric Colas

6.1.5 VASI_Cyto

**Keywords:** Data visualization, Biostatistics, Web Application

**Functional Description:** Web application for visualization of biologic data from datawarehouse of SISTM team.

**URL:** [https://gitlab.inria.fr/cneresta/VASICyto](https://gitlab.inria.fr/cneresta/VASICyto)

**Contact:** Clément Nerestan

**Partner:** INSERM
6.1.6 ccdf

Keyword: Biostatistics

**Functional Description:** Complex Hypothesis Testing Through Conditional Cumulative Distribution Function Estimation

**URL:** [https://CRAN.R-project.org/package=ccdf](https://CRAN.R-project.org/package=ccdf)

**Contact:** Boris Hejblum

7 New results

7.1 Methods for transcriptomic studies

The development of methods for the analysis of longitudinal gene expression data (encountered in many vaccine trials) started in 2014 and keeps being extended (Gauthier, bioRxiv 2021 [67], Villain, bioRxiv 2021 [70],). This work has started because none of the available approaches for longitudinal expression analysis was available to perform the analysis needed on the first VRI vaccine trial that was including repeated measures of gene expression. With the upcoming of single-cell RNAseq data, the need for adequate yet principled and sound methods for differential analysis of gene expression remains important. Our latest results leverage conditional to avoid making dubious distributional assumptions and accomodate the scRNAseq data specificities. Those methodologies have been applied to deepen our understanding of SARS-CoV-2 infection and identify potential biomarkers (Levy, iScience, 2021 [34]).

7.2 Translational vaccinology

7.2.1 Vaccine clinical trials

Main results of the EBL2002 Ebola vaccine trials have been published in two articles in Plos Medicine. Furthermore, the results of a therapeutic HIV vaccine trial have also been published in 2021 (Lèvy et al, J Virol 2021). A manuscript with the results of the Prevac Ebola vaccine trial has been submitted to the New England Journal of medicine.

Start of ANRS VRI06 first-in-human phase I trial of a novel vaccine concept targeting dendritic cells (here : as HIV vaccine); preparation of design of next generation Coronavirus vaccines based on that concept.

Furthermore, the ANRS VRI06 first-in-human phase I trial of a novel vaccine concept targeting dendritic cells (here as HIV vaccine) has started in 2021, and we have prepared the trial design of next generation Coronavirus vaccines based on that novel concept.

7.2.2 Covid treatment trials

The researchers of the team have been heavily mobilized the Coverage France clinical trial. Coverage France is a national multi-arm multi-stage (MAMS) adaptive trial platform for early treatments in Covid-infected outpatients in France that has been set-up in spring 2020 and obtained the “national priority” label (Capnet) (Duvignaud, Trials 2020). The trial was still ongoing in 2021 and has been adapted several times since its start, including evolution of the design and the evaluation of new treatment strategies. The trial has been stopped by the Scientific Advisory board in Dec 2021. A manuscript with the results of one of the tested treatment strategies (ciclesonide) has been submitted (co-first author: E Lhomme).

7.2.3 Systems immunology analyses

The systems vaccinology analyses of the Ebola vaccine trial EBL2001 within the Ebovac2 consortium has well advanced, a manuscript is being prepared. The systems vaccinology analyses of the ANRS VRI01 HIV vaccine trial are under revision in the Journal of Immunology.
7.2.4 Methodological developments for vaccine trials

The statistical methods developed by the team for the analysis of functional T-cell data using a bivariate modeling approach has been published in the Journal of Immunological Methods. We have strengthened our expertise in adaptive MAMS trial designs and their practical conduct thanks to the Coverage trial (see above). We have also started a new collaboration with Emilie Kaufmann (Scool team, Cristal, Inria Lille) in order to use bandit algorithms for vaccine clinical trial design and analyses. A master student will work on this topic in 2022.

Participants: Boris Hejblum, Edouard Lhomme, Mélanie Prague, Laura Richert, Rodolphe Thiébaut, Linda Wittkop.

8 Partnerships and cooperations

8.1 International initiatives

8.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program

DYNAMHIC

Title: DYNAMical modeling of HIV Cures

Duration: 2019 ->

Coordinator: Alison HILL (alhill@fas.harvard.edu)

Partners:

- Harvard University

Inria contact: Melanie Prague

Summary: Worldwide, over 35 million people are currently infected with HIV, and 2.3 million individuals are newly infected each year. While combination antiretroviral therapy can suppress virus replication, improving life expectancy and quality, it cannot eradicate the virus. A latent reservoir of virus exists in long-lived lymphocytes and can re-initiate the infection ("rebound") whenever treatment is stopped. Consequently, current therapy must be taken for life, and new research efforts are underway to find a permanent cure for HIV. Two general approaches are being taken to prevent HIV rebound and hence allow therapy to be completely stopped ("cure"). One approach, often called a "sterilizing cure", is to purge all the remaining latent virus (so called Berlin patient, Mississippi baby). Another approach, often called a “functional cure”, is to instead equip the immune system with the ability to control virus that reactivates from latency (so called elite controllers or post-treatment controllers). This provides the proof-of-concept for immunologically-mediated control. The studies we will analyze in this program are part of a larger effort to use therapies that enhance the immune response (known as "immunotherapy" and "therapeutic vaccine") to induce viral control either by clearing latent virus, boosting anti-viral immune responses, or both. The goal of this collaboration is to investigate the effect of each component of the potentially-curative intervention (vaccine and immunotherapy) and evaluate their synergy. We use statistical inference methods (developed by SISTM team) applied to mathematical models describing the dynamics of virus and immune cells (developed by Harvard team). Such approaches are superior to regression-based methods that ignore the underlying biological mechanisms. A Bayesian framework which estimates parameters at a population level allows us to explicitly test for differences between treatment groups, and, can overcome parameter identifiability issues that arise in complex, high-dimensional models. Indeed, these model can both provide mechanistic understanding of the biological process when administrating a treat- ment and help predicting the outcome for a new study. Previous studies have demonstrated that mechanistic models can accelerate the development of new drugs by
facilitating in silico simulation of clinical trials (e.g. for Hepatitis C antivirals). We aim at providing such a tool for the development of immunotherapies and vaccines for HIV cure.

8.1.2 Participation in other International Programs

NIPAH (Chine) – Scientific cooperation program France/Chine. (2019-2022). M. Prague is workpackage co-PI - Sino-French Agreement Aviesan. Sept. 2018 – Aug. 2023, 150,000 euros. To raise the challenge caused by Nipah virus we propose to develop a program that shall led to a better understanding of the epidemiology of the virus as well as the associated physiopathology. To develop new tools in the field of diagnosis, treatment and prevention of the infection. This grant aims at funding a 2 years of postdoc, travel and equipment expenses.

8.2 International research visitors

8.2.1 Visits of international scientists

Jane Heffernan

Status: Professor of mathematical modelling

Institution of origin: York University of Toronto

Country: Canada

Dates: from Sept 2020 to Aug 2021

Mobility program/type of mobility: research stay

8.3 European initiatives

8.3.1 FP7 and H2020 projects

IP-CURE-B: Immune profiling to guide host-directed interventions to cure HBV infections. Co-ordinated by Inserm (France), the project includes a total of 13 Beneficiaries: Centre Hospitalier Universitaire Vaudois (Switzerland), Karolinska Institutet (Sweden), Institut Pasteur (France), Universita degli studi di Parma (Italy), Fondazione IRCCS CA’ Granda – Ospedale maggiore policlinico (Italy), Universitaetsklinikum Freiburg (Germany), Ethniko Kai Kapodistriako Panepistimio Athi-non (Greece), Fundacio Hospital Universitari vall d’Hebron (Spain), Gilead Sciences Inc. (USA), Spring Bank Pharmaceuticals, Inc (USA), European Liver Patients Association (Belgium), Inserm Transfert SA (France). L Richert. Duration: 60 months 01/01/20-31/12/24. 409 632 Euros.


8.3.2 Other european programs/initiatives

EBOVAC2 (EBOVAC2): Development of a Prophylactic Ebola Vaccine Using a 2-Dose Heterologous Vaccination Regimen: Phase 2. Coordinated by Rodolphe Thiébaut with the following partners: Inserm (France), Labex VRI (France), Janssen Pharmaceutical Companies of Johnson and Johnson, London School of Hygiene and Tropical Medicine (United Kingdom), The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom), le Centre Muraz (Burkina Faso), Inserm Transfert (France). Duration: 72 months. 01/12/2014 - 30/11/2020. Total amount: IMI2 22,790,820 euros + EFPIA 50 710 893 €. Amount for SISTM: 2,930,196 euros.

EBOVAC1: Development of a Prophylactic Ebola Vaccine Using an Heterologous Prime-Boost Regimen. Coordinated by London School of Hygiene and Tropical Medicine (United Kingdom). Other beneficiaries: Janssen a Pharmaceutical Companies of Johnson and Johnson, The Chancellor, Masters and Scholars of
the University of Oxford (United Kingdom), Inserm (France), University of Sierra Leone (Sierra Leone), R. Thiébaut. Duration: 84 months. 01/12/2014 - 30/11/2021. 552,050 Euros.

EBOVAC3: Bringing a prophylactic Ebola vaccine to licensure. Coordinated by the London School of Hygiene and Tropical Medicine (United Kingdom). Other beneficiaries: Janssen a Pharmaceutical Companies of Johnson and Johnson, Inserm (France), The University of Antwerpen (Belgium), University of Sierra Leone (Sierra Leone), R. Thiébaut. Duration: 60 months. 01/06/2018 - 30/05/2023. 351,274 Euros.

PREVAC-UP: The Partnership for Research on Ebola VACcinations-extended follow-UP and clinical research capacity build-UP. SISTM is also involved in PREVAC-UP an EDCTP2 project in direct link with the research carried out on the Ebola vaccines. Coordinated by Inserm (France). Other beneficiaries: CNFRSR (Guinea), CERFIG (Guinea), LSHTM (UK), COMAHS (Sierra-Leone), NIAID (USA), NPHIL (Liberia), USTTB (Mali), Centre pour le Développement des Vaccins (Mali), Inserm Transfert SA (France), R. Thiébaut. Duration: 60 months. 01/01/2019 - 31/12/2023. 328,000 Euros.

CARE: Corona Accelerated R and D in Europe is an IMI2 funded project coordinated by Inserm which gathers 36 globally renowned academic institutions, pharmaceutical companies and non-profit research organisations which have committed to rapidly and efficiently address the COVID-19 emergent health threat. This major initiative aims at addressing two key objectives: the development of therapeutics to provide an emergency response towards the current COVID-19 pandemic and the development of therapeutics to address the current and/or future coronavirus outbreaks. To address both goals, the CARE consortium has carefully designed a comprehensive research and development (R and D) program around thoughtfully designed Target Product Profiles (TPP) of the urgently needed anticoVID-19 drugs. This includes small and large molecule discovery and Phase 1 and 2 clinical trials centred around three main pillars: drug repositioning, small-molecule drug discovery, and virus neutralising antibody discovery. These pillars reflect a bifocal strategy where efforts are geared towards (a) a rapid response against current COVID-19 pandemic and (b) a longer-term preparedness strategy against future coronavirus outbreaks. This will maximize the screening landscape of relevant therapeutic avenues and ensure effective therapeutics can be rapidly identified, pre-clinically tested and optimised for clinical-grade manufacturing and clinical testing. In this project, SISTM and EUCLID are working closely together with the support of the CREDIM in the WP5, W7 and WP8 with the respective objectives of providing statistical analysis and data modelling of the immune assays carried out in the project, bring some expert support to the clinical work and develop a LabKey-based platform for the integration and management of the data. Duration: 60 months. 01/04/2020 - 30/03/2025. 1,256,003 Euros.

ASCENT: Acceleration of Novel Coronavirus Serological Test Development and Seroprevalence Study: An African-European Initiative. ASCENT is an EDCTP2 projects involving 7 partners (Inserm, CHUV, EuroVacc, Utrecht University, Centre Muraz, SAMRC and CERFIG) from 6 different countries in Africa and Europe which will aim at assessing the real prevalence of the infection, the projection of the immunity acquired by the populations, and the evaluation of measures aimed to break the transmission in Africa. To do so ASCENT will implement in Burkina Faso, South Africa and Guinea, a novel robust and reproducible luminex-based serological diagnostic test with high throughput, sensitivity, specificity and rapid turnaround time. In this project, SISTM will be involved in statistical analysis of the tests data and will lead the WP3 which aims at modeling the epidemics. Duration: 24 months. 01/05/2020 - 30/04/2022. 37,500 Euros.

8.4 National initiatives

- Labex Vaccine Research Institute (VRI): There are strong collaborations with immunologists involved in the Labex Vaccine Research Institute (VRI) as Rodolphe Thiébaut and Laura Richert are leading the Data science division (197,095 euros in 2020) VRI.

- RHU SHIVA: since November 2019, R. Thiébaut is collaborating in the SHIVA RHU to work on the Integration and systems biology of MRI-cSmall Vessel Disease biomarkers. The budget for the SISTM team corresponds to the costs of a postdoc position. The duration of the RHU SHIVA is 60 months.

8.4.1 Expert Appraisals

- Rodolphe Thiébaut was a member of the CNU 46.04 (Biostatistiques, informatique médicale et technologies de communication) until Sept 2021.
- Rodolphe Thiébaut is a member of the Scientific Council of Inserm.
- Edouard Lhomme is an expert for ANRS-MIE (French Agency for Research on AIDS and Viral Hepatitis) in the COVID19 outpatient working group. He coordinates the response to the ministerial referral on the development of ambulatory clinical research.
- Edouard Lhomme is a member of the independent committee of international trials ANRS 177 DUETTO evaluating innovative antiretroviral therapy for people living with HIV and the French vaccine COVIBOOST trial, evaluating different vaccine boost strategy against COVID-19.
- L Richert is member of the independent committee of several national clinical trials (out of which one Covid antibody treatment phase I trial; and 1 Covid vaccine phase I trials).
- L. Richert is an expert reviewer for the PHRC funding programme.
- L. Richert is a member of the CNU 46.04 (Biostatistiques, informatique médicale et technologies de communication).
- Linda Wittkop is member of the external ethics and scientific advisory board of the EU-funded project VACCELERATE.
- Linda Wittkop is member of the CESREES (Comité éthique et scientifique pour les Recherches, les Etudes et les Evaluation dans le domaine de la santé).

8.4.2 Various Partnership

The project team members are involved in:

- F-CRIN (French clinical research infrastructure network), initiated in 2012 by ANR under "Programme des Investissements d’avenir". (L Richert and E Lhomme).
- INCA (Institut National du Cancer) funded the project Evaluation de l’efficacité d’un traitement sur l’évolution de la taille tumurale et autres critères de survie : développement de modèles conjoints. (Principal PI Virginie Rondeau Inserm U1219, Mélanie Prague is responsible of Work package 4 “mechanistic modeling of cancer: 5800 euros”).
- Contrat Initiation ANRS MoDeL-CI: Modeling the HIV epidemic in Ivory Coast (Principal PI Eric Ouattara Inserm U1219 in collaboration with University College London, Mélanie Prague is listed as a collaborator).
- TARPON (Traitement Automatique des Résumés de Passages aux urgences pour un Observatoire National), laureate project from the 2nd Health Data Hub calls for projects, great challenge “Improving medical diagnostics through Artificial Intelligence” and Bpifrance. (Principal PI E. Lagarde Inserm U1219 in collaboration with University Hospital of Bordeaux. Marta Avalos is listed as a collaborator).
• CESIR IV (Combination of Studies on Health and Road Safety - 4th project) funded by ONISR DSR. (Principal PI E. Lagarde Inserm U1219. Marta Avalos is listed as a collaborator).

• Collaboration with Inserm PRC (pôle Recherche clinique).

• Collaboration with Inserm RECap (Recherche en Epidémiologie Clinique et en Santé Publique) network.

Participants:  Marta Avalos, Boris Hejblum, Edouard Lhomme, Mélanie Prague, Laura Richert, Rodolphe Thiébaut, Linda Wittkop.

9 Dissemination

9.1 Promoting scientific activities

9.1.1 Scientific events: organisation


Boris Hejblum organizes the fortnightly joint Biostatistics Seminar Series between the Bordeaux Population Health Inserm Research Center and the Bordeaux University Public Health Department.

General chair, scientific chair.

Rodolphe Thiébaut is a member of the scientific committee of the IWHOD International Workshop on HIV Observational Databases since 2013 (IWHOD) and chair in 2020-2021.

9.1.2 Scientific events: selection

Member of the conference program committees.

Boris Hejblum was a member of the Scientific Program Committee of the 42nd Conference of the International Society for Clinical Biostatistics (ISCB) in Lyon (France).

Marta Avalos was a member of the Program Committee of the ACM Conference on Health, Inference, and Learning, 2021 and the Machine Learning for Health – ML4H Workshop, 2021.

Laura Richert was a member of the Scientific Programm Committee of the 15ème conférence francophone d’EPIdémiologie CLINique (EPICLIN 2021).

Communication in conference.

M. Prague, M. Lavielle, Online, SAMBA: A novel method for fast automatic model building in nonlinear mixed-effects models 5th Virus dynamics workshop, online, 4-6 oct. 2021.

Alexandre M., Prague M., Thiébaut R. - Dynamics of the humoral immune response to a two-dose heterologous vaccine regimen against Ebola virus Online, PAGE Conference 2-7 Sept 2021.


Alexandre M, Thiébaut R, Prague M Accounting for time-dependant confounding variables in mechanistic ODE model: simulations and application to a vaccine trial, Online, ISCB 19/22 July. 2021.

9.1.3 Journal

Member of the editorial boards.

Melanie Prague is associate editor of "International journal of Biostatistics" (since 2018).
Reviewer - reviewing activities.
Bayesian Analysis, PLOS Computational Biology, WIREs: Data Mining And Knowledge Discovery, Statistical Applications in Genetics and Molecular Biology, International Journal of Biostatistics (B Hejblum).
IMIA Yearb Med Inform (M Avalos)

9.1.4 Invited talks
R Thiébaut: "Definition of surrogate marker by mechanistic modeling of the vaccine response applied to SARS-Cov-2" Journées GDR "Statistiques et santé", 21-22 October 2021, Virtual.
R Thiébaut 24th Neuromuscular Days - September 9-10, 2021, Marseille, Artificial intelligence: what will it change for tomorrow's health?
L. Richert: Eccmid 2021 congress (July 9-12, 2021), in the symposium "Update on new vaccines: Immunogenicity, safety and recent progress with vector-based vaccines against Filoviruses".
R. Genuer Paris, 08/10/2021 SAMM seminar, Paris 1 University.
R. Genuer Vannes10/12/2021 Vannes statistics meeting.
M. Avalos “A decision-making tool to adjust for abnormal levels in complete blood count tests”. Biomathematical Networks Workshop, 28th Oct 2021, Virtual/Valparaiso, Chile.
B. Hejblum, 8th Channel Network Conference 8 April 2021— Leveraging random effects to estimate the impact of NPIs on epidemic dynamics across French regions.
B. Hejblum, IA en Nouvelle Aquitaine (PIIA) 28 juin 2021— Reconnaissance automatique des populations cellulaires.
B. Hejblum, Journée Reproductibilité de la Recherche de l’INRAE 5 février 2021—Introduction à la reproductibilité pour la recherche en biométrie – définitions, enjeux et illustrations.
B. Hejblum, SMPGD 25 January 2021 — Distribution-free complex hypothesis testing for single-cell RNA-seq differential expression analysis.
B. Hejblum, workshop StatOmique, Paris (France) — Inférence de proportions cellulaires à partir de données RNA-seq avec CIBERSORT dans un essai clinique de suivi de patients COVID.

9.1.5 Scientific expertise
Rodolphe Thiébaut is an expert for INCA (Institut National du Cancer) for the PHRC (Programme hospitalier de recherche Clinique en cancérologie) and for the PRME (Programme de recherche médico-économique en cancérologie).
Rodolphe Thiébaut is a member of the committee “Biologie des Systèmes et Cancer (Plan Cancer)”, a member of the Scientific Advisory Board of the “Institut Pierre Louis d’Épidémiologie et de Santé Publique” (UPMC, Dir: Dominique Costagliola), a member of the independent committee of international trials ODYSSEY and SMILE, a member of the scientific council of Muraz’s Center (Bobo-Dioulasso, Burkina Faso).

Swiss National Science Foundation: Application reviewer in 2021 (2 projets) - M. Prague.
Agence Nationale de Recherche sur le SIDA et hépatites virales / ANRS MIE: Application reviewer for the CSS13 “Recherches cliniques” - M Prague (since 2019) and L Wittkop.

Boris Hejblum was a reviewer for PhD (CIFRE) application at the Agence Nationale Recherche Technologie (ANRT).

9.1.6 Research administration

L Wittkop: Coordination of the ANRS MIE working group “Reinforcement of vaccine clinical trial platforms”.

Boris Hejblum is a member of the chairing committee of the Société Française de Biométrie, the French Chapter of the International Biometric Society.

Boris Hejblum is a board member of the “Machine Learning et Intelligence Artificielle” (MALIA) group of the French Society of Statistics (SfDs).

Boris Hejblum is the correspondant for the French Biometric Society to the Channel Network region of the International Biometrics Society.

Marta Avalos is general secretary of the “Statistics and Sport” group of SfDs.

Rodolphe Thiébaut is the director of the department of Public Health in University of Bordeaux and a member of the Inserm Scientific Council.

Laura Richert is coordinator of the Clinical epidemiology module of the Clinical Investigations Center (CIC1401 Bordeaux).

Mélanie Prague is co-president of the communication group of French statistical society since 2018, in charge of redefining de condition of sponsoring SfdS by enterprises

Mélanie Prague is a member of the Bureau “Action Coordonnée Modélisation” ANRS MIE.

Mélanie Prague is a member of the "Commission de Développement Technologique" since 2016. Evaluation and expertise of grants applications to Inria for the transfer of technology, partnership with industries and development of new technological tools.

Mélanie Prague is a member of the "Commission Emploi-recherche" since 2017. Participation in the scientific committee for attribution of research grants for PhD students and Postdoctoral fellows.

9.2 Teaching - Supervision - Juries

9.2.1 Teaching

- In class teaching
  - Master: Rodolphe Thiébaut is head of the Digital Public Health graduate program, University of Bordeaux.
  - Master: All the permanent members and several PhD students teaches in the Master of Public Health (M1 Santé publique, M2 Biostatistique and/or M2 Epidemiology) and the Digital Public Health graduate program, University of Bordeaux.
  - Master: Marta Avalos teaches in the Master of Applied Mathematics and Statistics (1st and/or 2nd year), University of Bordeaux.
  - Master: Marta Avalos teaches in the 2nd year of the Master of "Management international : Développement pharmaceutique, Production et Qualité opérationnelle”, University of Bordeaux.
  - Bachelor: Laura Richert, Linda Wittkop and Edouard Lhomme teach in PACES and DFASM1-3 for Medical degree at Univ. Bordeaux.
  - Master: Edouard Lhomme teaches in the Master of Vaccinology from basic immunology to social sciences of health (University Paris-Est Créteil, UPEC).
– Teaching unit coordination: Laura Richert, Linda Wittkop, Rodolphe Thiébaut, Robin Genuer, Boris Hejblum and Marta Avalos coordinate several teaching units of Master in Public Health (Biostatistics, Epidemiology, Public Health). Laura Richert coordinates the teaching unit "Experimental Designs" (M2 Epidemiology) and the teaching unit "critical article reading" (across 4 years of medical school), University of Bordeaux.

– Summer school: Mélanie Prague teaches in the summer school, University of Bordeaux.

– Boris Hejblum teaches a 3-day graduate course “Bayesian analysis for biomedical research” at the University of Copenhaguen.

– Mélanie Prague teaches "Missing Data" at ENSAI Master Level.

• E-learning

– Master: Marta Avalos is head of the first year of the e-learning program of the Master of Public Health, University of Bordeaux.

– Master: Marta Avalos teaches in the e-learning program of the Master of Public Health (1st and 2nd year).

– ODL University Course: Robin Genuer is head of the Diplôme universitaire "Méthodes statistiques en santé". Mélanie Prague teaches in the Diplôme universitaire "Méthodes statistiques de régression en épidémiologie".

– ODL University Course: Edouard Lhomme co-coordinates and teaches in the Diplôme universitaire "Recherche Clinique".

9.2.2 Supervision


• Marta Avalos supervised PhD student: Madelyn Rojas (University of Bordeaux). Home, Leisure, and Sports Injuries in France : Study of the risk factors associated with home injuries in the MAVIE cohort (Injury Epidemiology team, Inserm U1219, ED SP2) co-directed with Emmanuel Lagarde (Inserm).

• PhD in progress: Iris Ganser, Evaluation of event-based internet biosurveillance for multi-regional detection of seasonal influenza onset, co-directed by David Buckeridge (McGill University) and Rodolphe Thiébaut, from Oct 2020.

• PhD in progress: Benjamin Hivert, Hierarchical modeling for integrative analysis of high-dimensional, high-throughput, multi-modal cell and molecular data for immunology research, co-directed by Boris Hejblum and Rodolphe Thiébaut, from Oct 2020.

• PhD Ipsen, CIFRE in progress: Hélène Savel "Statistical analysis of OMICS data for the treatment response prediction in early clinical development in a context of the generation of virtual patients to run In Silico Clinical Trials", directed by Laura Richert, from Oct 2020.

• PhD in progress: Marie Alexandre "Mechanistic modeling and optimization of vaccine response in HIV and Ebola", co-directed by Mélanie Prague and Rodolphe Thiébaut, from Oct 2018.

• PhD in progress: Iris Ganser, Evaluation of event-based internet biosurveillance for multi-regional detection of seasonal influenza onset, co-directed by David Buckeridge (McGill University) and Rodolphe Thiébaut, from Oct 2020.
Marta Avalos supervised Master 2 internship students: Clémence Métayer (co-supervised with Laurence Delhaes, Inserm U1045) and Alexandre Naprous (co-supervised with Emmanuel Lagarde, Inserm U1219.

Boris Hejblum supervised Emma Avisou first year Master internship "Analysis of the relationship between antibody levels and efficacy of different vaccines against the Covid-19 epidemic.

Boris Hejblum supervised Clément Bonnet first year Master internship "Simulation of epidemiological curves of COVID-19 with a SEIRAH model integrated on an interactive web application with shiny and golem".

Mélanie Huchon "Variable selection methods for high dimensional survival data" M2 internship from 01/03/2021 to 31/08/2021, supervised by R Genuer.


Paul Hermabessiere –Master 2 (L Wittkop) Association of Metabolic Associated Fatty Liver Disease (MAFLD) with morbidity and mortality in chronic viral hepatitis B treated in the ANRS CO22 HEPATHER cohort.

Gautier Boillet (with M Chalouni and L Wittkop) Comparison of the predictive performance of PAGE-B, FIB-4 and FibroScan scores for the occurrence of hepatocellular carcinoma in chronically hepatitis B infected patients in the ANRS CO22 HEPATHER cohort.


9.2.3 Juries

Robin Genuer HDR defense 12/01/21 "Contributions to random forest methods for various data analysis problems".

HDR Jury : Mickael White, Robin Genuer (R Thiébaut).

University thesis jury : Durier, Maryan Morel, Aurélie Bannay, Emma Gerard (R Thiébaut).

Jury for the practice thesis : Julien Coelho (R Thiébaut)

Jury recruitment iSTARS tenure, Lisbon (R Thiébaut).

University thesis jury: Linda Wittkop was rapporteur of Valérie Potard (Sorbonne University).

Linda Wittkop: Follow-up committee of Houreratou Barry, Iris Ganser.

R Genuer, rapporteur of the University thesis of Rico Blaser, defended on 07/12/2021 at the London School of Economics.


B Hejblum: University thesis jury of Shayma Bel Hechmi (Université Paris-Saclay, 12/07/2021).

Marta Avalos was involved in the PhD defense jury of Madelyn Rojas (University of Bordeaux) as PhD supervisor. Defense date: 1st June.

Marta Avalos is a member of the follow-up dissertation committee of 1 PhD student: Alexandre Conanec (Statistics, IMB, ED MI). Defense date: 10 November.
• Edouard Lhomme, Laura Richert and Linda Wittkop participated to the juries of medical thesis defenses, Medical School of Bordeaux University.

• Marta Avalos participated to the recruitment jury of the Master of "Management international : Développement pharmaceutique, Production et Qualité opérationnelle", University of Bordeaux.

9.3 Popularization

9.3.1 Articles and contents

Inria 2021 podcast: What is the purpose of digital health research? Inria Podcasts.

9.3.2 Education

M Avalos: Presentation in the “Salon Culture et Jeux Mathématiques” (May 27-30). Our talk (with C Derquenne and B Gelein, representing the SFdS group “Stat et Sport”) was titled "Du collège à l’université : la Science et le Sport font équipe.” An article was also written for the review of this show.


9.3.3 Interventions

Media interventions:

• Le monde 25/09/2021 Covid-19: despite five weeks of continuous decline in metropolitan France, epidemiologists remain cautious.

• La tribune 19/7/2021 Artificial intelligence and epidemiology: two keys for public health?

• France 2, JT 20h, 01/09/2021 Chronique Damien Mascret France TV Info.

• Clues as to why women are more resistant to COVID-19. Le Parisien 06/01/2021 Covid-19.


• Epidemiologists confused by the evolution of the epidemic. Le Monde 20/02/21.

• The rise of variants under high surveillance. Le Monde 10/04/21.

• Although exposed, teachers "protect themselves well" from Covid19. Le Parisien 04/29/2021.

• Health restrictions lifted. Horizon 05/05/2021.

• Five things to know about: Mixing and matching coronavirus vaccines . Le Monde 23/07/2021.

• "Scientists divided on the possibility of removing the mask”. Le Monde 29/07/2021.

• "The race of vaccination in front of the fourth wave”. Le Parisien 06/09/2021.

• "Is there really a correlation between the incidence rate and vaccination?”. La voix du Nord 4/11/2021.


• "Covid-19 outbreak flares up again in Europe: update” Update.

Participants: Marta Avalos, Robin Genuer, Boris Hejblum, Edouard Lhomme, Mélanie Prague, Laura Richert, Rodolphe Thiébaut, Linda Wittkop.
10 Scientific production

10.1 Major publications


10.2 Publications of the year

International journals


URL: [https://hal.inria.fr/hal-03481499](https://hal.inria.fr/hal-03481499).


URL: [https://hal.inria.fr/hal-03477973](https://hal.inria.fr/hal-03477973).


URL: [https://hal.inria.fr/hal-03100447](https://hal.inria.fr/hal-03100447).

H. Lorenzo, O. Cloarec, R. Thiébaut and J. Saracco. ‘Data-Driven Sparse Partial Least Squares’. In: *Statistical Analysis and Data Mining* (25th Dec. 2021). URL: [https://hal.inria.fr/hal-03368956](https://hal.inria.fr/hal-03368956).


URL: [https://hal.inria.fr/hal-03534044](https://hal.inria.fr/hal-03534044).


URL: [https://hal.inria.fr/hal-03534057](https://hal.inria.fr/hal-03534057).


URL: [https://hal.inria.fr/hal-03481526](https://hal.inria.fr/hal-03481526).


URL: [https://hal.archives-ouvertes.fr/hal-03189015](https://hal.archives-ouvertes.fr/hal-03189015).


URL: [https://hal.inria.fr/hal-03186351](https://hal.inria.fr/hal-03186351).


URL: [https://hal.archives-ouvertes.fr/hal-03136348](https://hal.archives-ouvertes.fr/hal-03136348).


Conferences without proceedings


Edition (books, proceedings, special issue of a journal)


Doctoral dissertations and habilitation theses


Reports & preprints


10.3 Other

Patents


10.4 Cited publications


