



IN PARTNERSHIP WITH:
INSERM

Université de Bordeaux

Activity Report 2018

Project-Team SISTM

Statistics In System biology and Translational
Medicine

RESEARCH CENTER
Bordeaux - Sud-Ouest

THEME
**Modeling and Control for Life Sci-
ences**

Table of contents

1. Team, Visitors, External Collaborators	1
2. Overall Objectives	2
2.1. Overall Theoretical Objectives	2
2.2. Overall Applied Objectives	3
3. Research Program	3
3.1. Mechanistic modelling	3
3.2. High dimensional data	3
4. Application Domains	4
4.1. Systems Biology and Translational medicine	4
4.2. The case of HIV immunology	4
4.3. The case of Ebola vaccine development	5
5. Highlights of the Year	5
6. New Software and Platforms	6
6.1. marqLevAlg	6
6.2. VSURF	6
6.3. NPflow	6
6.4. COVVSURF	6
6.5. clogitLasso	6
6.6. TcGSA	7
6.7. NIMROD	7
6.8. tcgsaseq	7
6.9. cytometree	7
6.10. CRTgeeDR	7
6.11. ludic	8
6.12. CoDaPCA	8
6.13. Left-censored Lasso	8
6.14. dd-sPLS	8
6.15. kernscr	8
6.16. phenotypr	9
6.17. R2GUESS	9
7. New Results	9
7.1. Mecanistic modeling	9
7.1.1. Methodology	9
7.1.2. Applications	9
7.1.2.1. Applied to IL-7 therapy	9
7.1.2.2. Applied to development of new HIV immunotherapies and vaccines	10
7.1.2.3. Applied to development of Ebola vaccines	10
7.2. Statistical learning methods for high-dimensional data	10
7.2.1. Automatic analysis of cell populations	10
7.2.2. Missing Value Treatment in Longitudinal High Dimensional Supervised Problems	10
7.2.3. Left-censored data treatment in High Dimensional Supervised Problems	10
7.3. Analysis of results from Clinical trials and cohorts	11
7.3.1. In the HIV field	11
7.3.2. In the Ebola field	11
7.3.3. In cluster randomized trials	11
7.4. Conferences	11
8. Bilateral Contracts and Grants with Industry	12
9. Partnerships and Cooperations	13
9.1. Regional Initiatives	13

9.2. National Initiatives	13
9.2.1. Expert Appraisals	13
9.2.2. Various Partnership	13
9.3. European Initiatives	14
9.3.1. FP7 & H2020 Projects	14
9.3.2. Collaborations in European Programs, Except FP7 & H2020	14
9.3.3. Collaborations with Major European Organizations	15
9.4. International Initiatives	15
9.4.1. Inria International Labs	15
9.4.2. Inria International Partners	16
9.5. International Research Visitors	16
10. Dissemination	17
10.1. Promoting Scientific Activities	17
10.1.1. Scientific Events Organisation	17
10.1.2. Scientific Events Selection	17
10.1.2.1. Chair of Conference Program Committees	17
10.1.2.2. Member of the Conference Program Committees	17
10.1.3. Journal	18
10.1.3.1. Member of the Editorial Boards	18
10.1.3.2. Reviewer - Reviewing Activities	18
10.1.4. Invited Talks	18
10.1.5. Leadership within the Scientific Community	18
10.1.6. Scientific Expertise	19
10.1.7. Research Administration	19
10.2. Education	19
10.2.1. Teaching	19
10.2.2. Supervision	20
10.2.3. Juries	21
10.3. Popularization	21
10.3.1. Internal or external Inria responsibilities	21
10.3.2. Articles and contents	22
10.3.3. Interventions	22
10.3.4. Internal action	22
11. Bibliography	22

Project-Team SISTM

Creation of the Team: 2013 April 02, updated into Project-Team: 2015 January 01

Keywords:

Computer Science and Digital Science:

- 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

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. - Pharmacokinetics and dynamics
- B2.4.2. - Drug resistance

1. Team, Visitors, External Collaborators

Research Scientists

- Daniel Commenges [INSERM, Senior Researcher]
- Mélanie Prague [Inria, Researcher]

Faculty Members

- Rodolphe Thiébaud [Team leader, Univ de Bordeaux, Professor, HDR]
- Marta Avalos Fernandez [Univ de Bordeaux, Associate Professor, HDR]
- Robin Genuer [Univ de Bordeaux, Associate Professor]
- Boris Hejblum [Univ de Bordeaux, Associate Professor]
- Laura Richert [Univ de Bordeaux, Associate Professor]

Post-Doctoral Fellows

- Irene Balelli [INSERM]
- Binbin Xu [Univ de Bordeaux]

PhD Students

- Marie Alexandre [Inria, from Oct 2018]

Louis Capitaine [Univ de Bordeaux]
Marine Gauthier [Univ de Bordeaux, from Oct 2018]
Edouard Lhomme [Univ de Bordeaux]
Hadrien Lorenzo [INSERM]
Chloé Pasin [Univ de Bordeaux, until Sep 2018]
Perrine Soret [INSERM, until Jun 2018]
Laura Villain [Univ de Bordeaux, until Sep 2018]

Technical staff

Chariff Alkhasim [INSERM, until Apr 2018]
Fabien Barbier [INSERM]
Solenne Delahaye [INSERM]
Melany Durand [INSERM, from Mar 2018]
Myriam Lopez [INSERM, until Feb 2018]
Van Hung Tran [INSERM, from Nov 2018]

Interns

Roxane Coueron [INSERM, from Mar 2018 until Sep 2018]
Anthony Devaux [Univ de Bordeaux, from Apr 2018 until Jun 2018]
Marine Gauthier [Inria, from Feb 2018 until Jul 2018]
Julien Rouar [Inria, from Feb 2018 until Aug 2018]

Administrative Assistants

Cecile Boutros [Inria, from Sep 2018]
Nathalie Robin [Inria, until Aug 2018]
Sandrine Darmigny [Inserm, Assistant]
Anton Ottavi [Inserm, European Projects coordinator]
Destandau Eugénie [Inserm, Communication]

2. Overall Objectives

2.1. Overall Theoretical Objectives

The overall objective of SISTM is to develop statistical methods for the integrative analysis of health data, especially those related to clinical immunology to answer specific questions risen in the application field. To reach this objective we are developing statistical methods belonging to two main research areas:

- Statistical and mechanistic modeling, especially based on ordinary differential equation systems, fitted to population and sparse data
- Statistical learning methods in the context of high-dimensional data

These two approaches are used for addressing different types of questions. Statistical learning methods are developed and applied to deal with the high dimensional characteristics of the data. The outcome of this research leads to hypotheses linked to a restricted number of markers. Mechanistic models are then developed and used for modeling the dynamics of a few markers. For example, regularized methods can be used to select relevant genes among 20000 measured with microarray/RNA-seq technologies, whereas differential equations can be used to capture the dynamics and relationship between several genes followed over time by a q-PCR assay or RNA-seq.

2.2. Overall Applied Objectives

Data are generated in clinical trials or biological experimentations. Our main application of interest is the immune response to vaccine or other immune interventions (such as exogenous cytokines), mainly in the context of HIV infection. The methods developed in this context can be applied in other circumstances but the focus of the team on immunology is important for the relevance of the results and their translation into practice, thanks to a longstanding collaboration with several immunologists and the implication of the team in the Labex Vaccine Research Institute (<http://vaccine-research-institute.fr>). Examples of objectives related to this application field are:

- To understand how immune response is generated with immune interventions (vaccines or interleukines)
- To predict what would be the immune response to a given immune intervention for designing next studies and adapting interventions to individual patients

3. Research Program

3.1. Mechanistic modelling

When studying the dynamics of a given marker, say the HIV concentration in the blood (HIV viral load), one can for instance use descriptive models summarising the dynamics over time in term of slopes of the trajectories [47]. These slopes can be compared between treatment groups or according to patients' characteristics. Another way for analysing these data is to define a mathematical model based on the biological knowledge of what drives HIV dynamics. In this case, it is mainly the availability of target cells (the CD4+ T lymphocytes), the production and death rates of infected cells and the clearance of the viral particles that impact the dynamics. Then, a mathematical model most often based on ordinary differential equations (ODE) can be written [40]. Estimating the parameters of this model to fit observed HIV viral load gave a crucial insight in HIV pathogenesis as it revealed the very short half-life of the virions and infected cells and therefore a very high turnover of the virus, making mutations a very frequent event [39].

Having a good mechanistic model in a biomedical context such as HIV infection opens doors to various applications beyond a good understanding of the data. Global and individual predictions can be excellent because of the external validity of a model based on main biological mechanisms. Control theory may serve for defining optimal interventions or optimal designs to evaluate new interventions [32]. Finally, these models can capture explicitly the complex relationship between several processes that change over time and may therefore challenge other proposed approaches such as marginal structural models to deal with causal associations in epidemiology [31].

Therefore, we postulate that this type of model could be very useful in the context of our research that is in complex biological systems. The definition of the model needs to identify the parameter values that fit the data. In clinical research this is challenging because data are sparse, and often unbalanced, coming from populations of subjects. A substantial inter-individual variability is always present and needs to be accounted as this is the main source of information. Although many approaches have been developed to estimate the parameters of non-linear mixed models [43], [50], [35], [41], [36], [49], the difficulty associated with the complexity of ODE models and the sparsity of the data leading to identifiability issues need further research.

3.2. High dimensional data

With the availability of omics data such as genomics (DNA), transcriptomics (RNA) or proteomics (proteins), but also other types of data, such as those arising from the combination of large observational databases (e.g. in pharmacoepidemiology or environmental epidemiology), high-dimensional data have become increasingly common. Use of molecular biological technics such as Polymerase Chain Reaction (PCR) allows for amplification of DNA or RNA sequences. Nowadays, microarray and Next Generation Sequencing (NGS) techniques

give the possibility to explore very large portions of the genome. Furthermore, other assays have also evolved, and traditional measures such as cytometry or imaging have become new sources of big data. Therefore, in the context of HIV research, the dimension of the datasets has much grown in term of number of variables per individual than in term of number of included patients although this latter is also growing thanks to the multi-cohort collaborations such as CASCADE or COHERE organized in the EuroCoord network¹. As an example, in a recent phase 1/2 clinical trial evaluating the safety and the immunological response to a dendritic cell-based HIV vaccine, 19 infected patients were included. Bringing together data on cell count, cytokine production, gene expression and viral genome change led to a 20 Go database [46]. This is far from big databases faced in other areas but constitutes a revolution in clinical research where clinical trials of hundred of patients sized few hundred of Ko at most. Therefore, more than the storage and calculation capacities, the challenge is the comprehensive analysis of these datasets.

The objective is either to select the relevant information or to summarize it for understanding or prediction purposes. When dealing with high dimensional data, the methodological challenge arises from the fact that datasets typically contain many variables, much more than observations. Hence, multiple testing is an obvious issue that needs to be taken into account [44]. Furthermore, conventional methods, such as linear models, are inefficient and most of the time even inapplicable. Specific methods have been developed, often derived from the machine learning field, such as regularization methods [48]. The integrative analysis of large datasets is challenging. For instance, one may want to look at the correlation between two large scale matrices composed by the transcriptome in the one hand and the proteome on the other hand [37]. The comprehensive analysis of these large datasets concerning several levels from molecular pathways to clinical response of a population of patients needs specific approaches and a very close collaboration with the providers of data that is the immunologists, the virologists, the clinicians...

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" [45], "Systems vaccinology" [42], "Systems medicine" [34]. From the statistician 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. The case of HIV immunology

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 [38]. 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) [33]. However, the implementation of such strategy is difficult especially in developing countries. Some HIV infected individuals do not tolerate antiretroviral regimen or did not

¹ see online at <http://www.eurocoord.net>

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. The case of Ebola vaccine development

In response to the recent outbreak of Ebola virus disease in West Africa, the clinical development of some candidate to Ebola vaccine has been accelerated. Several vectors, mostly encoding glycoprotein of the virus, were tested in Phase I-II studies in order to assess their safety and immunogenicity. One of the main question of interest there is the antibody response induced by vaccination, as some non-human primates studies have shown protection against the virus when antibody levels were high enough. Although bridging studies still have to be developed, antibodies are thus considered as a criterium of interest. The challenge is then to evaluate the durability of the antibody response, whether it be at an individual or population level, in order to evaluate the impact of a vaccine strategy in case of an epidemic. Moreover, we are interested in the factors associated to this antibody response, and even more the other immune markers (from both innate and adaptative immune response) able to predict antibody levels. As those relationship are non-linear, sophisticated statistical and mathematical methods are developed in order to address these questions. A systems medicine approach using multidimensional immunogenicity data from clinical trials and statistical models can help to understand vaccine mechanisms and improve the selection of optimised vaccine strategies for clinical trials.

5. Highlights of the Year

5.1. Highlights of the Year

- Launch of the Graduate's School Digital Public Health (PI: R Thiebaut) including the Master of Public Health Data Sciences
- Launch of the IMI project EBOVAC3 in which R Thiebaut is leader of the workpackage "Modelling". Concomitantly, we have obtained the first results of the modelling of the response to the Ebola vaccine developed with Janssen company (submitted to Journal of Virology).
- A new step of the work on IL-7 therapy in HIV infected patients has been achieved through the optimization of the administration of the injections. Approaches from statistical modelling and control theory demonstrated the feasibility of reducing the administration of IL-7 while improving its efficacy.
- The project on the automatic recognition of cell populations through high dimensional cytometry data has reached a successful stage with two important publications. It is now applied to clinical trial datasets available through the Vaccine Research Insitute.
- The data warehouse system developed through the EHVA European consortium is settled in its version 1.0 and will be used for the storage of all SISTM datasets as well as to implement the software developed for the analysis of immunological data.
- Funding of the EDCTP Prevac-UP in which M Prague is leader of the workpackage "System vaccinology approach". The aim is to develop an integrative analysis of all immunological data generated to understand antibodies response to Ebola vaccination.
- Funding of the Franco-Sino INSERM project on NiPAH virus in which M Prague is leader of the workpackage "Modeling, biostatistics and bioinformatics". The aim of this workpackage is to conduct state of the art quantitative analyses of effects of therapeutic and vaccine strategies, as well as providing a framework to bridge results from in vitro to in vivo and between different animal models.

- Marta Avalos undertook a 6-month research visit to Data61 (CSIRO, Canberra, Australia) in 2017. This collaboration has reached a successful stage with one publication in NeurIPS 2018.

6. New Software and Platforms

6.1. marqLevAlg

KEYWORDS: Optimization - Biostatistics

FUNCTIONAL DESCRIPTION: An R package for function optimization. Available on CRAN, this package performs a minimization of function based on the Marquardt-Levenberg algorithm. This package is really useful when the surface to optimize is non-strictly convex or far from a quadratic function. A new convergence criterion, the relative distance to maximum (RDM), allows the user to have a better confidence in the stopping points, other than basic algorithm stabilization.

- Contact: Melanie Prague
- URL: <https://cran.r-project.org/web/packages/marqLevAlg/index.html>

6.2. VSURF

Variable Selection Using Random Forests

KEYWORDS: Classification - Statistics - Machine learning - Regression

FUNCTIONAL DESCRIPTION: An R package for Variable Selection Using Random Forests. Available on CRAN, this package performs an automatic (meaning completely data-driven) variable selection procedure. Originally designed to deal with high dimensional data, it can also be applied to standard datasets.

- Contact: Robin Genuer
- URL: <https://github.com/robingenuer/VSURF>

6.3. NPflow

Bayesian Nonparametrics for Automatic Gating of Flow-Cytometry Data

KEYWORDS: Bayesian estimation - Bioinformatics - Biostatistics

FUNCTIONAL DESCRIPTION: Dirichlet process mixture of multivariate normal, skew normal or skew t-distributions modeling oriented towards flow-cytometry data pre-processing applications.

- Contact: Boris Hejblum
- URL: <https://cran.r-project.org/web/packages/NPflow/>

6.4. COVVSURF

Combination of Clustering Of Variables and Variable Selection Using Random Forests

KEYWORDS: Classification - Statistics - Cluster - Machine learning - Regression

- Contact: Robin Genuer
- URL: <https://github.com/robingenuer/CoVVSURF>

6.5. clogitLasso

KEYWORDS: Biostatistics - Bioinformatics - Machine learning - Regression

FUNCTIONAL DESCRIPTION: R package to fit a sequence of conditional logistic regression models with lasso, for small to large sized samples.

RELEASE FUNCTIONAL DESCRIPTION: Optimisation

- Partner: DRUGS-SAFE
- Contact: Marta Avalos Fernandez
- URL: <https://cran.r-project.org/web/packages/clogitLasso/index.html>

6.6. TcGSA

Time-course Gene Set Analysis

KEYWORDS: Bioinformatics - Genomics

FUNCTIONAL DESCRIPTION: An R package for the gene set analysis of longitudinal gene expression data sets. This package implements a Time-course Gene Set Analysis method and provides useful plotting functions facilitating the interpretation of the results.

- Contact: Boris Hejblum
- URL: <https://cran.r-project.org/web/packages/TcGSA/index.html>

6.7. NIMROD

Normal approximation Inference in Models with Random effects based on Ordinary Differential equations

KEYWORDS: Ordinary differential equations - Statistical modeling

FUNCTIONAL DESCRIPTION: We have written a specific program called NIMROD for estimating parameter of ODE based population models.

- Contact: Melanie Prague
- URL: <http://etudes.isped.u-bordeaux2.fr/BIOSTATISTIQUE/NIMROD/documentation/html/index.html>

6.8. tcgsaseq

Time-Course Gene Set Analysis for RNA-Seq Data

KEYWORDS: Genomics - Biostatistics - Statistical modeling - RNA-seq - Gene Set Analysis

FUNCTIONAL DESCRIPTION: Gene set analysis of longitudinal RNA-seq data with variance component score test accounting for data heteroscedasticity through precision weights.

- Contact: Boris Hejblum
- URL: <https://cran.r-project.org/web/packages/tcgsaseq/index.html>

6.9. cytometree

KEYWORDS: Clustering - Biostatistics - Bioinformatics

FUNCTIONAL DESCRIPTION: Given the hypothesis of a bimodal distribution of cells for each marker, the algorithm constructs a binary tree, the nodes of which are subpopulations of cells. At each node, observed cells and markers are modeled by both a family of normal distributions and a family of bimodal normal mixture distributions. Splitting is done according to a normalized difference of AIC between the two families.

- Contact: Boris Hejblum
- URL: <https://cran.r-project.org/web/packages/cytometree/index.html>

6.10. CRTgeeDR

KEYWORDS: Missing data - Statistics - Regression

FUNCTIONAL DESCRIPTION: The CRTgeeDR package allows you to estimate parameters in a regression model (with possibly a link function). It allows treatment augmentation and IPW for missing outcome. It is particularly of use when the goal is to estimate the intervention effect of a prevention strategy against epidemics in cluster randomised trials.

- Contact: Melanie Prague
- URL: <https://cran.r-project.org/web/packages/CRTgeeDR/index.html>

6.11. ludic

KEYWORDS: Probability - Biostatistics

FUNCTIONAL DESCRIPTION: An R package to perform probabilistic record Linkage Using only Diagnosis Codes without direct identifiers, using C++ code to speed up computations. Available on CRAN, development version on github.

- Contact: Boris Hejblum
- URL: <https://cran.r-project.org/web/packages/ludic/index.html>

6.12. CoDaPCA

KEYWORDS: Unsupervised learning - PCA

FUNCTIONAL DESCRIPTION: R functions associated to the article Avalos et al. Representation Learning of Compositional Data. NeurIPS 2018 <http://papers.nips.cc/paper/7902-representation-learning-of-compositional-data>

- Contact: Marta Avalos Fernandez
- URL: <https://github.com/sistm/CoDa-PCA>

6.13. Left-censored Lasso

KEYWORDS: Biostatistics - Machine learning

FUNCTIONAL DESCRIPTION: R function associated to the article Soret et al. Lasso regularization for left-censored Gaussian outcome and high-dimensional predictors. BMC Medical Research Methodology (2018) 18:159 <https://doi.org/10.1186/s12874-018-0609-4>

RELEASE FUNCTIONAL DESCRIPTION: <https://github.com/psBiostat/left-censored-Lasso>

- Contact: Marta Avalos Fernandez

6.14. dd-sPLS

Data-Driven Sparse PLS

KEYWORDS: Marker selection - Classification - Regression - Missing data - Multi-Block - High Dimensional Data - PLS - SVD

FUNCTIONAL DESCRIPTION: Allows to build Multi-Data-Driven Sparse PLS models. Multi-blocks with high-dimensional settings are particularly sensible to this. What's more it deals with missing samples (entire lines missing per block) thanks to the Koh-Lanta algorithm. SVD decompositions permit to offer a fast and controlled method.

- Contact: Hadrien Lorenzo
- URL: <https://hadrienlorenzo.netlify.com/projects/ddspls/>

6.15. kernscr

KEYWORDS: Genomics - Biostatistics

FUNCTIONAL DESCRIPTION: An R package to perform KERNel machine score test for pathway analysis in the presence of Semi-Competing Risks

- Contact: Boris Hejblum
- URL: <https://CRAN.R-project.org/package=kernscr>

6.16. phenotypr

KEYWORDS: Phenotyping - Automatic labelling - Automatic Learning

FUNCTIONAL DESCRIPTION: Machine learning prediction algorithm for predicting a clinical phenotype from structured diagnostic data and CUI occurrence data collected from medical reports, previously processed by NLP approaches.

- Contact: Boris Hejblum
- URL: <https://github.com/borishejblum/phenotypr>

6.17. R2GUESS

Graphical processing Unit Evolutionary Stochastic Search

FUNCTIONAL DESCRIPTION: R2GUESS package is a wrapper of the GUESS (Graphical processing Unit Evolutionary Stochastic Search) program. GUESS is a computationally optimised C++ implementation of a fully Bayesian variable selection approach that can analyse, in a genome-wide context, single and multiple responses in an integrated way. The program uses packages from the GNU Scientific Library (GSL) and offers the possibility to re-route computationally intensive linear algebra operations towards the Graphical Processing Unit (GPU) through the use of proprietary CULA-dense library.

- Contact: Rodolphe Thiebaut

7. New Results

7.1. Mechanistic modeling

7.1.1. Methodology

Estimation methods in mechanistic models can be seen as an inverse problem, in which we want to recover the individual parameters that have produced some observations. In collaboration with the Inria MONC & M3DISIM team, we propose a method for estimation in ODE with mixed effects on parameters based on Kalman based filter (also known as linear quadratic estimation (LQE)) that consist in correcting the original dynamic at each time by a feedback control.

7.1.2. Applications

7.1.2.1. Applied to IL-7 therapy

We have developed two approaches to optimize the injection of IL-7 in HIV-infected patients: one based on a statistical model and one based on Piecewise Deterministic Markov Processes (PDMP).

Villain L, Commenges D, Pasin C, Prague M, Thiébaud R. Adaptive protocols based on predictions from a mechanistic model of the effect of IL7 on CD4 counts. *Statistics in Medicine*. 2018;38:221-235.

Pasin C, Dufour F, Villain L, Zhang H, Thiébaud R. Controlling IL-7 Injections in HIV-Infected Patients. *Bulletin of Mathematical Biology*. 2018;80:2349-2377.

7.1.2.2. *Applied to development of new HIV immunotherapies and vaccines*

HIV infection can be treated but not cured with combination antiretroviral therapy, and new therapies that instead target the host immune response to infection are now being developed. Two recent studies of such immunotherapies, conducted in an animal model (SIV-infected rhesus macaques), have shown that agents which target the innate immune receptor TLR7 along with recombining viral-vector vaccines can prevent or control the rebound in viremia that usually accompanies the discontinuation of antiretroviral drugs. However, the mechanism of action of these therapies remains unknown. In collaboration with Harvard School of public health and Harvard program for evolutionary dynamics, we delineate the best model and the best procedure for treatment effect selection in order to delineate the each of each immunotherapies and design subsequent trials.

7.1.2.3. *Applied to development of Ebola vaccines*

We have analysed data from three clinical trials conducted under the EBOVAC1 consortium in 4 different countries (UK, Kenya and Uganda/Tanzania) and assessing the safety and immunogenicity of prime-boost vaccine regimens combining one adenovirus-based vector and one modified vaccine Ankara. In particular, we have modeled the dynamics of the humoral immune response following the boost immunization and fitted both linear mixed models and ODEs-based mechanistic models to the antibody concentrations data. This analysis allowed the estimation of the durability of the antibody response, as well as the identification of factors of variability of the response.

Pasin C, Balelli I, Van Effelterre T, Bockstal V, Soloforosi L, Prague M, Douoguih M, Thiébaud R. Dynamics of the humoral immune response to a prime-boost Ebola vaccine: quantification and sources of variation. Under revision in *Journal of Virology*.

7.2. Statistical learning methods for high-dimensional data

7.2.1. *Automatic analysis of cell populations*

We have developed two different approaches to classify the cell populations according to the high dimensional data obtained with flow cytometry assays. The first one is based on an interesting development of Dirichlet processes and the second one is based on a simple tree classification providing very high performances:

Hejblum BP, Alkhassim C, Gottardo R, Caron F, Thiébaud R. Sequential Dirichlet process mixtures of multivariate skew t-distributions for model-based clustering of flow cytometry data. *Annals of Applied Statistics*. In press.

Commenges D, Alkhassim C, Gottardo R, Hejblum B, Thiébaud R. cytometree: A binary tree algorithm for automatic gating in cytometry analysis. *Cytometry A*. 2018;93:1132-1140.

7.2.2. *Missing Value Treatment in Longitudinal High Dimensional Supervised Problems*

Poor blood sample quality introduces a large number of missing values in the context of sequencing data production. Furthermore, strong technical biases may force the analyst to remove the considered sequenced samples. Then entire day dependent data are then missing.

We have developed a regularized SVD based method using the temporal structure (through multi-block approach) of the missing values to estimate missing values with the objective of predicting uni-variate or multivariate regression responses but also classification problems. That regularizing method uses soft-thresholding on the co-variance matrices implying natural variable selection of covariate and response through a single hyper-parameter to be tuned.

7.2.3. *Left-censored data treatment in High Dimensional Supervised Problems*

Data could be censored either by the limit of detection or the limit of quantification. We have developed a regularized method for handling high-dimensional exposure data in the presence of censored values in the field of HIV that could be applied to other fields.

Soret P, Avalos M, Wittkop L, Commenges D, Thiébaud R. Lasso regularization for left-censored Gaussian outcome and high-dimensional predictors. *BMC Med Res Methodol*. 2018 Dec 4;18(1):159.

7.3. Analysis of results from Clinical trials and cohorts

7.3.1. In the HIV field

We have performed the statistical analyses of the immunogenicity endpoints, including high dimensional assays such as gene expression (RNA Seq), of two HIV vaccine clinical trials: 1) ANRS VRI01, a randomized phase I/II trial evaluating for different prime boost vaccine strategies in healthy volunteers; 2) ANRS 159 LIGHT, a randomized phase II trial comparing a prime-boost therapeutic HIV vaccine strategy to placebo in HIV-infected patients undergoing antiretroviral treatment interruption. The results of each of these two trials have been presented as an oral presentation at the HIV R4P conference in Madrid in October 2018 (Richert L et al. and Lacabaratz C et al). Integrative statistical analyses using sPLS methods (as developed by the team) are currently ongoing to relate markers from different high-dimensional immunogenicity assays or virological assays to each other.

7.3.2. In the Ebola field

We have performed a review of all existing clinical trials available to evaluate Ebola vaccines in macaques and humans.

Gross L, Lhomme E, Pasin C, Richert L, Thiébaud R. Ebola vaccine development: Systematic review of pre-clinical and clinical studies, and meta-analysis of determinants of antibody response variability after vaccination. *International Journal of Infectious Diseases*. 2018. pii: S1201-9712(18)34457-6.

7.3.3. In cluster randomized trials

Accounting for missing outcome is highly important to recover unbiased results of treatment effects. Weighting approaches are less common compared to multilevel multiple imputation to analyse clustered data with missing outcome. >In collaboration with Duke University, we compared the two approaches and evaluated their performances to conclude that weighted approach should be considered a viable strategy to account for missing outcomes in cluster randomized trials

In cluster randomized trials, it is often desirable to improve the understanding of intervention effects in the presence of dissemination/spillover. In collaboration with Rhode Island university and Yale University, we aim at proposing innovative approach to analyze the TasP ANRS 12249 trial. We proposed innovative methods to individual, disseminated and overall effect of a clustered intervention based on counterfactual averages in the presence of dissemination.

7.4. Conferences

Members of the team were involved in several talks during conferences and colloquium.

- Big Data and Information Analytics 2018 BigDIA Conference – 17-19th December 2018, Houston, Texas, USA – Random Forests for high-dimensional longitudinal data [invited talk]
- Big Data and Information Analytics 2018 BigDIA Conference – 17-19th December 2018, Houston, Texas, USA – Analysis of high-dimensional longitudinal data from the French health-administrative databases using machine learning methods: performance comparison between LSTM neural networks and Lasso for the analysis of the risk of road traffic crashes associated with medicinal drug consumption [invited talk]
- Journées Recherche et Santé (JRS) Inserm Phénotypage Clinique et biologie des systèmes – 22 Novembre 2018, Institut Imagine Paris, France - Visualisation des omics et des données cliniques [invited talk]
- Center for Modelling and Simulation in the Biosciences (BIOMS) Symposium 2018 – 1-2nd Octobre 2018, Heidelberg, Germany - Finding the cells in the middle of the data [invited talk]

- McGill University, Department of Epidemiology Seminar – 24th September 2018, Montreal, Canada - Big Data In Vaccine Clinical Trials: A Dive Into Data Science [invited talk]
- 4th Neurepiomics summer school - 17-20th September 2018, Arcachon, France - Systems biology approaches applied to omics data [invited talk]
- International Workshop in honor of Daniel Commenges' 70th birthday- June 4-5, 2018, Bordeaux, France - The mechanistic model point of view of causality [invited talk]
- Population Approach Group Europe (PAGE meeting) - May 29 - June 1, 2018, Montreux, Switzerland - Use of mathematical modeling for optimizing and adapting immunotherapy protocols in HIV-infected patients [oral contribution]
- International Symposium on HIV and Emerging Infectious Diseases (ISHEID) - May 16-18, 2018, Marseille, France - In silico clinical research, keep it dynamic! [invited talk]
- International Conference « Statistics and Health » - 11-12 January 2018, Toulouse Institute of Mathematics, Toulouse - Use of mathematical modeling for accelerating and personalizing clinical trials [invited talk]
- Workshop Developments in cluster randomised and stepped wedge designs, London, 21-22 Nov. 2018 Performance of weighting as an alternative to multilevel multiple imputation in cluster randomized trials with missing binary outcomes [oral contribution]
- International Biometrics Society, Barcelona, Spain, 09-13 July 2018. Optimizing the administration of IL7. [oral contribution]
- International Biometrics Society, Barcelona, Spain, 09-13 July 2018. Fitting pharmacokinetics data with a population-based Kalman filters [oral contribution]
- International Biometrics Society, Barcelona, Spain, 09-13 July 2018. Random Forests for high-dimensional longitudinal data. [oral contribution]
- ENBIS European Network for Business and Industrial Statistics conference, 2-6 Sept 2018. Mechanistic modeling for in silico trials [oral contribution]
- IMI 10th Anniversary Scientific Symposium, Brussels, Belgium, 22-23 October 2018. Modelling the humoral immune response to Ebola vaccine [oral contribution]
- Montreal University, Faculté de Pharmacie, Department of Mathematical pharmacology Seminar – 21th December 2018, Montreal, Canada - Mechanistic modeling for in silico trials [invited talk]

8. Bilateral Contracts and Grants with Industry

8.1. Contracts and Grants with Industry

Implication in research for the development of Ebola vaccine has lead to several indirect contracts with industry:

- The EBOVAC1, EBOVAC2 and EBOVAC3 project, collaboration with Janssen from Johnson et Johnson.
- The Prevac trial vaccine trial leads to collaboration with Merck and Janssen. The purpose of this study is to evaluate the safety and immunogenicity of three vaccine strategies that may prevent Ebola virus disease (EVD) events in children and adults. Participants will receive either the Ad26.ZEBOV (rHAd26) vaccine with a MVA-BN-Filo (MVA) boost, or the rVSV Δ G-ZEBOV-GP (rVSV) vaccine with or without boosting, or placebo. The Prevac-UP project is set as a continuation of Prevac trial in the same framework.

9. Partnerships and Cooperations

9.1. Regional Initiatives

The team have strong links with :

- Research teams of the research center INSERM U1219 : "Injury Epidemiology, Transport, Occupation" (IETO), "Biostatistics", "Pharmacoepidemiology and population impact of drugs", "Multi-morbidity and public health in patients with HIV or Hepatitis" (MORPH3Eus), "Computer research applied to health" (ERIAS) emerging research team.
- Bordeaux and Limoges CHU ("Centre Hospitalier Universitaire").
- Institut Bergonié, Univ Bordeaux through the Euclid platform
- Inria Project-team MONC, M3DISIM and CQFD

The project team members are involved in:

- EUCLID/F-CRIN clinical trials platform (Laura Richert)
- The research project "Self-management of injury risk and decision support systems based on predictive computer modelling. Development, implementation and evaluation in the MAVIE cohort study" funded by the Nouvelle-Aquitaine regional council (Marta Avalos).
- Phenotyping from Electronic Health Records pilot project in cooperation with with the ERIAS Inserm emerging team in Bordeaux and the Rheumatology service from the Bordeaux Hospital (Boris Hejblum)
- A cancer research project in collaboration with Inria MONC team on glioblastoma

9.2. National Initiatives

- Labex Vaccine Research Institute (VRI) There are strong collaborations with immunologists involved in the Labex Vaccine Research Institute (VRI) as Rodolphe Thiébaud is leading the Data science division (previously Biostatistics/Bioinformatics) <http://vaccine-research-institute.fr>.
- Collaboration with Inserm PRC (pôle Recherche clinique).
- Collaboration with Inserm Reacting (REsearch and ACTION targeting emerging infectious diseases) network

9.2.1. Expert Appraisals

- Rodolphe Thiébaud is a member of the CNU 46.04 (Biostatistiques, informatique médicale et technologies de communication).
- Rodolphe Thiébaud is a member of the Scientific Council of INSERM.
- Mélanie Prague is an expert for ANRS (France Recherche Nord&Sud Sida-HIV Hépatites) in the CSS 3 (Recherches cliniques et physiopathologiques dans l'infection à VIH) and AC 47 (Dynamique et contrôle des épidémies VIH et hépatites).
- Laura Richert is an expert for the PHRC (Programme hospitalier de recherche Clinique).
- Marta Avalos is an expert for the ANSM (Agence nationale de sécurité du médicament et des produits de santé)

9.2.2. Various Partnership

The project team members are involved in:

- DRUGS-SAFE platform funded by ANSM (Marta Avalos). Initiated in 2015-2018. Renewed for 2019.
- F-CRIN (French clinical research infrastructure network) was initiated in 2012 by ANR under two sources of founding "INBS/Infrastructures nationales en biologie et en santé" and "Programme des Investissements d'avenir". (Laura Richert)

- I-REIVAC is the French vaccine research network. This network is part of the “Consortium de Recherche en Vaccinologie (CoReVac)” created by the “Institut de Microbiologie et des Maladies Infectieuses (IMMI)”. (Laura Richert)
- INCA (Institut National du Cancer) funded the project « Evaluation de l’efficacité d’un traitement sur l’évolution de la taille tumorale 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).

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

The member of SISTM Team are involved in EHVA (European HIV Vaccine Alliance):

Program: Most information about this program can be found at <http://www.ehv-a.eu>.

Coordinator: Rodolphe Thiébaud is Work Package leader of the WP10 "Data Integration".

Other partners: The EHVA encompasses 39 partners, each with the expertise to promote a comprehensive approach to the development of an effective HIV vaccine. The international alliance, which includes academic and industrial research partners from all over Europe, as well as sub-Saharan Africa and North America, will work to discover and progress novel vaccine candidates through the clinic.

Abstract: With 37 million people living with HIV worldwide, and over 2 million new infections diagnosed each year, an effective vaccine is regarded as the most potent public health strategy for addressing the pandemic. Despite the many advances in the understanding, treatment and prevention of HIV made over the past 30 years, the development of broadly-effective HIV vaccine has remained unachievable. EHVA plans to develop and implement:

Discovery Platform with the goal of generating novel vaccine candidates inducing potent neutralizing and non-neutralizing antibody responses and T-cell responses

Immune Profiling Platform with the goal of ranking novel and existing (benchmark) vaccine candidates on the basis of the immune profile

Data Management/Integration/Down-Selection Platform, with the goal of providing statistical tools for the analysis and interpretation of complex data and algorithms for the efficient selection of vaccines

Clinical Trials Platform with the goal of accelerating the clinical development of novel vaccines and the early prediction of vaccine failure.

9.3.2. Collaborations in European Programs, Except FP7 & H2020

Program: The EBOVAC2 project is one of 8 projects funded under IMI Ebola+ programme that was launched in response to the Ebola virus disease outbreak. The project aims to assess the safety and efficacy of a novel prime boost preventive vaccine regimen against Ebola Virus Disease (EVD). Project acronym: EBOVAC2 Project title: EBOVAC2 Coordinator: Rodolphe Thiébaud Other partners: Inserm (France), Labex VRI (France), Janssen Pharmaceutical Companies of Johnson & Johnson, London School of Hygiene & Tropical Medicine (United Kingdom), The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom), Le Centre Muraz (Burkina Faso), Inserm Transfert (France) Given the urgent need for an preventive Ebola vaccine strategy in the context of the current epidemic, the clinical development plan follows an expedited scheme, aiming at starting a Phase 2B large scale safety and immunogenicity study as soon as possible while assuring the safety of the trial participants.

Phase 1 trials to assess the safety and immunogenicity data of the candidate prime-boost regimen in healthy volunteers are ongoing in the UK, the US and Kenya and Uganda. A further study site has been approved to start in Tanzania. Both prime-boost combinations (Ad26.ZEBOV prime + MVA-BN-Filo boost; and MVA-BN-Filo prime + Ad26.ZEBOV boost) administered at different intervals are being tested in these trials.

Phase 2 trials (this project) are planned to start as soon as the post-prime safety and immunogenicity data from the UK Phase I are available. Phase 2 trials will be conducted in healthy volunteers in Europe (France and UK) and non-epidemic African countries (to be determined). HIV positive adults will also be vaccinated in African countries. The rationale for inclusion of European volunteers in Phase 2, in addition to the trials in Africa, is to allow for higher sensitivity in safety signal detection in populations with low incidence of febrile illnesses, to generate negative control specimens for assay development, to allow for inclusion of health care workers or military personnel that may be deployed to Ebola-endemic regions.

EBOVAC3 As a follow-up project, the IMI-2 funded EBOVAC3 project, has started in 2018. In this project, the vaccine strategy will be further evaluated in specific populations in Africa (infants in Guinea and Sierra Leone; and front line workers in RDC). The project includes a work package on modelling, which is led by Rodolphe Thiebaut.

9.3.3. Collaborations with Major European Organizations

University of Oxford;

London School of Hygiene and Tropical Medicine;

University Hospital Hamburg;

Heinrich Pette Institute for Experimental Virology, Hambourg;

MRC, University College London;

MRC Biostatistics Unit, University of Cambridge.

9.4. International Initiatives

- Prevac-Up: this EDCTP-funded research program will collect and analyze important information on the effectiveness and maintainability of Ebola vaccination in sub-Saharan Africa.
- NIPAH program: a Franco-Chinese collaboration of 1 million euros per country over 5 years to develop a program to better understand Nipah Virus infection, its physiopathology and to develop new tools for diagnosis, treatment and prevention.

9.4.1. Inria International Labs

Inria@SiliconValley

Associate Team involved in the International Lab:

9.4.1.1. SWAGR

Title: Statistical Workforce for Advanced Genomics using RNAseq

International Partner (Institution - Laboratory - Researcher):

RAND Corporation (United States) - Statistics group - Denis Agniel

Start year: 2018

See also: <https://team.inria.fr/swagr/>

The SWAGR Associate Team aims at bringing together a statistical workforce for advanced genomics using RNAseq. SWAGR combines the biostatistics experience of the SISTM team from Inria BSO with the mathematical expertise of the statistics group at the RAND Corporation in an effort to improve RNAseq data analysis methods by developing a flexible, robust, and mathematically principled framework for detecting differential gene expression. Gene expression, measured through the RNAseq technology, has the potential of revealing deep and complex biological mechanisms underlying human health. However, there is currently a critical limitation in widely adopted approaches for the analysis of such data, as edgeR, DESeq2 and limma-voom can all be shown to fail to control the type-I error, leading to an inflation of false positives in analysis results. False positives are an important issue in all of science. In particular in biomedical research when costly studies are failing to reproduce earlier results, this is a pressing issue. SWAGR propose to develop a rigorous statistical framework modeling complex transcriptomic studies using RNAseq by leveraging the synergies between the works of B. Hejblum and D. Agniel. The new method will be implemented in open-source software as a Bioconductor R package, and a user friendly web-application will be made available to help dissemination. The new method will be applied to clinical studies to yield significant biological results, in particular in vaccine trials through existing SISTM partnerships. The developed method is anticipated to become a new standard for the analysis of RNAseq data, which are rapidly becoming common in biomedical studies, and has therefore the potential for a large impact.

9.4.2. Inria International Partners

Fred Hutchinson Cancer center, Seattle;

Baylor Institute for Immunology (Dallas);

Duke University -Duke Global Health Institute, Elizabeth Turner.

Harvard University - Department of Population Medicine - Rui Wang. M. Prague is a co-PI in an amfAR project for co-supervision of a PhD student.

university of San Diego UCSD / Harvard School of public health - NIH program project grant "Revealing Reservoirs During Rebound", Harvard School of Public Health (HSPH) and the University of California, San Diego (P01AI131385, total budget \$1.5M/yr for 5 years starting Oct 2017, both university manage the funding. Mélanie Prague is part of modelling unit of the "Quantitative Methods" research project (budget \$220,000/yr). The principal investigator for this core is Victor de Grutolla (HSPH) The overall goal of this grant is to characterize viral rebound following antiretroviral therapy cessation in cohorts of patients who have started therapy early in infection, as well as in a cohort of terminally-ill patients who will interrupt therapy before death and subsequently donate their bodies to research.

Harvard University - program for evolutionary dynamics - Alison Hill and Martin Nowak. Project submitted by M Prague for the Inria associated team with this laboratory.

Collaborations through clinical trials: NIH and University of Minnesota for the Prevac trial, NGO Alima for the Prevac trial, Several African clinical sites for Ebovac2 and Prevac trials;

Tianxi Cai from Harvard University on developing methods for the linkage and analysis of Electronic Health Records data (Boris Hejblum).

Katherine Liao from Harvard University on the analysis of Electronic Health Records data in the context of Rheumatoid Arthritis (Boris Hejblum).

Sylvia Richardson from MRC Biostatistics Unit Cambridge University on the scaling up of nonparametric Bayesian approaches to Big data (Boris Hejblum).

Machine learning team Data61 at CSIRO, Australia (Marta Avalos)

9.5. International Research Visitors

9.5.1. Visits to International Teams

Marta Avalos visited Marcela Henríquez Henríquez 1 week in December 2018, Medical School, Pontifical Catholic University of Chile (Chile). Travel grant from the French Embassy in Chile and the French Institute of Chile

Boris Hejblum visited the RAND Corporation for a week in July 2018 for a research collaboration with Denis Agniel (Santa-Monica, USA).

Boris Hejblum visited the MRC BSU for 10 days in October 2018 for a research collaboration with Sylvia Richardson & Paul Kirk.

Chloé Pasin visited Harvard School of Public Health on May 9-11 2018 and Columbia University, Departement of Pathology and Cell Biology on May 14-17 2018.

Mélanie Prague got invited in University of Montreal (Canada) for a 3-days research trip in the pharmacy department on December 19-22 2018 to be member of the jury committee of the PhD defense of Steven Sanche (Effet des antiretroviraux sur la pathogénèse du VIH), as well as giving an invited talk.

Mélanie Prague spent 3 weeks in Boston as an invited researcher in Harvard School of Public Health & Department of evolutionary dynamics, Biostatistics department in January and february 2018.

9.5.1.1. Research Stays Abroad

- Laura Richert spent 6 months (March-August 2018) as a visting researcher at the Heinrich Pette Institut for Experimental Virology, Research Department Virus Immunology, Hamburg, Germany (head: Marcus Altfeld).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Rodolphe Thiébaud, Boris Hejblum and Marta Avalos co-organized an invited session “Predicting health outcomes from administrative claims data and electronic health records” in the 4th International Conference on Big Data and Information Analytics. Theories, Algorithms and Applications in Data Science (BigDIA). Dec 17-19, 2018, Houston (USA). The invited session was chair by Rodolphe Thiébaud.

Robin Genuer Co-organizes a reading group called Smiling in Bordeaux (<http://www.math.u-bordeaux.fr/~mchave100p/smiling/>)

Boris Hejblum organizes the Biostatistics Seminar Series at the Bordeaux Public Health Inserm Research Center

Mélanie Prague organized the "Déjeuners scientifiques" at the "Journées de la statistique française" 2018.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Mélanie Prague was the organizer of a day on "Use of mathematical models for personalized medicine" at the International Conference « Statistics and Health: personalized medicine » - 11-12 January 2018, Toulouse Institute of Mathematics, Toulouse France

10.1.2.2. Member of the Conference Program Committees

Daniel Commenges is a member of the scientific committee of the International Biometric Conference Barcelona, July 2018 (<http://2018.biometricconference.org>)

Mélanie Prague is a member of the scientific committee of CIMI conference “Statistics in Health - personalised medicine” (<http://www.cimi.univ-toulouse.fr/mib/en/conference-statistics-and-health>), Toulouse 2018, 10-12 January

Marta Avalos was a member of the program committee of the annual meeting of the francophone Machine Learning community, CAp 2018, Rouen, June 2018 (<http://cap2018.litislab.fr/Comites-en.html>)

Rodolphe Thiébaud was a member of the scientific committee of the national conference on clinical research (EPICLIN)

Rodolphe Thiébaud is a member of the scientific committee of the IWHOD International Workshop on HIV Observational Databases since 2013 (<http://newsite.iwhod.org/Committee>)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Lifetime Data Analysis (Daniel Commenges)

Statistics Surveys (Daniel Commenges)

Associate editor of International journal of Biostatistics (Melanie Prague)

10.1.3.2. Reviewer - Reviewing Activities

ADAC (Robin Genuer)

AIDS (Rodolphe Thiébaud)

Am J Public Health (Mélanie Prague)

Biostatistics (Laura Richert)

Biometrics (Mélanie Prague, Boris Hejblum)

IMIA Yearb Med Inform (Marta Avalos)

International Journal of Epidemiology (Daniel Commenges)

Journal of Computational and Graphical Statistics (Robin Genuer)

Journal of the Royal Statistical Society: Interaction (Mélanie Prague)

Journal of Statistical Computation and Simulation (Boris Hejblum)

JRSS-B (Mélanie Prague)

Scientific Reports (Laura Richert)

Society of clinical trial (Mélanie Prague)

Statistical Methods in Medical Research (Robin Genuer, Mélanie Prague)

Statistical science (Mélanie Prague)

Trials (Laura Richert)

10.1.4. Invited Talks

Mélanie Prague gave 4 invited talks in 2018 (Montreal university, Canada; Department of evolutionary dynamics Harvard Boston USA; Dracula Inria team in Lyon France; institute Gustave Roussy Biostatistics Group Paris).

Robin Genuer gave one invited talk at the "Séminaire de Probabilités et de Statistique du Laboratoire de Mathématiques Paul Painlevé de Lille".

Boris Hejblum gave 2 invited talks in 2018 (Genotoul Biostat/Bioinfo Day in Toulouse; Section of Biostatistics, University of Copenhagen, Denmark)

Marta Avalos gave 2 invited talks at the workshop “Big Data: la revolución de la información en la investigación biomédica”, 18-19th December 2018, Santiago de Chile (Chile)

10.1.5. Leadership within the Scientific Community

Rodolphe Thiébaud and Chloé Pasin are elected members of the “collège des écoles doctorales”, University of Bordeaux

Daniel Commenges is President of the French Region of the International Biometric Society

Mélanie Prague is an elected member of the “Young statistician group” of SFdS (French Society of Statistics)

Mélanie Prague is in charge of the group responsible for the communication of the SFdS - in charge of organizing the sponsoring of the society by public and private companies.

Laura Richert is a member of F-CRIN Steering Committee

Marta Avalos is general secretary of the “Statistics and Sport group” of SFdS (French Society of Statistics)

10.1.6. Scientific Expertise

- Rodolphe Thiébaud 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 Thiébaud is a member of the Membre du CNU 46.04 (Biostatistiques, informatique médicale et technologies de communication).
- Rodolphe Thiébaud is a member of the Scientific Council of INSERM.
- Rodolphe Thiébaud 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’Epidémiologie 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)
- Mélanie Prague is an expert for ANRS (France Recherche Nord&Sud Sida-HIV Hépatites) in the CSS 3 (Recherches cliniques et physiopathologiques dans l’infection à VIH) and AC 47.
- Laura Richert is an expert for the PHRC (Programme hospitalier de recherche Clinique).
- Marta Avalos is an expert for the ANSM (Agence nationale de sécurité du médicament et des produits de santé)

10.1.7. Research Administration

Daniel Commenges is the director of the Biostat-Info axis in the Inserm BPH (Bordeaux Public Health) institute.

Rodolphe Thiébaud is an elected member of the research committee (health sector) in University of Bordeaux and a member of the INSERM Scientific Council

10.2. Education

10.2.1. Teaching

In class teaching

Master : Robin Genuer, teaches in the two years of the Master of Public Health (M1 Santé publique, M2 Biostatistique) and 2nd year of the "Modélisation Stochastique et Statistique" Master, University of Bordeaux.

Master : Boris Hejblum, teaches in the two years of the Master of Public Health (M1 Santé publique, M2 Biostatistique).

Master : Rodolphe Thiébaud, teaches in the two years of the Master of Public Health, and he is head of the Epidemiology speciality of the second year of the Master of Public Health.

Master : Laura Richert teaches in the Master of Public Health at ISPED, Univ. Bordeaux, France (M2 Biostatistiques, M2 Epidémiologie).

Master : Mélanie Prague teaches in the Master of Public Health at ISPED, Univ. Bordeaux, France (M2 Biostatistiques).

Master : Marta Avalos teaches in the two years of the Master of Public Health (M1 Santé publique, M2 Biostatistique), the two years of the Master of Applied Mathematics and Statistics, and the 2nd year of the Master of "Management international : Développement pharmaceutique, Production et Qualité opérationnelle", Univ. of Bordeaux.

Master: Chloé Pasin, Laura Villain, Hadrien Lorenzo and Louis Capitaine are teaching assistants for the two years.

Edouard Lhomme teaches in the Master of Public Health at ISPED, Univ. Bordeaux (M2 Epidémiologie) and in the Master of Vaccinology from basic immunology to social sciences of health (University Paris-Est Créteil, UPEC)

Bachelor : Laura Richert teaches in PACES and DFASM1-3 for Medical degree at Univ. Bordeaux

Edouard Lhomme teaches in PACES and DFASM1-3 for Medical degree at Univ. Bordeaux

Bachelor: Mélanie Prague and Boris Hejblum teach in the third year ingenious school ENSAI, Rennes.

Summer School: the SISTM team member teach in the ISPED Summer school.

E-learning

Marta Avalos is head of the first year of the e-learning program of the Master of Public Health, and teaches in it.

Mélanie Prague teaches in the Diplôme universitaire "Méthodes statistiques de régression en épidémiologie".

Boris Hejblum teaches in the Diplôme universitaire "Méthodes statistiques en santé.

Laura Richert teaches in the Diplôme universitaire "Recherche Clinique".

Robin Genuer is head of the Diplôme universitaire "Méthodes statistiques en santé and participated to the IdEx Bordeaux University "Défi numérique" project "BeginR" (<http://beginr.u-bordeaux.fr>).

10.2.2. Supervision

Master internship: Marie Alexandre, *PKPD modeling in pre-clinical development*, co-directed by Mélanie Prague with Nicolas Frances Roche Basel Switzerland (01/04/2018 - 31/09/2018)

Master internship (M1): Anthony Devaux, *Gene expression analysis with the R software*, directed by Boris Hejblum (01/06/2018 - 31/08/2018)

Master internship (M2): Roxane Coueron, *Sample size estimation for a microbiome study*, co-directed by Boris Hejblum with Hélène Savel, CHU Bordeaux (01/02/2018 - 31/08/2018)

Master internship: Marine Gauthier, *Variance component test for RNA-seq data analysis*, directed by Boris Hejblum (01/02/2018 - 31/07/2018)

Master internship: Julien Rouar, *PCA for absolute and relative abundance microbiota data: survey and implementation of methods*, co-directed by Marta Avalos with Cheng Soon Ong and Richard Nock, Data61, Australia (26/02/2018 - 31/08/2018)

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

PhD in progress: Marine Gauthier "Methods for bulk and single-cell RNA-seq data analysis in vaccine research", co-directed by Boris Hejblum and Rodolphe Thiébaud, from Sept 2018.

PhD in progress : Soufiane Ajana "Comparison of linear and non-linear machine learning approaches to predict Age-related Macular Disease (AMD) risk in a survival framework", co-supervised by Boris Hejblum and Hélène Jacquemin-Gadda (Inserm) and Cécile Delcourt (Inserm), from Sept 2016.

PhD in progress: Perrine Soret, *Modélisation de données longitudinales en grande dimension*, from Oct 2014, directed by Marta Avalos.

PhD in progress : Wenjia Wang "Modèle de Rasch", co-directed by Daniel Commenges with Mickael Guedj CIFRE Pharnext, from Oct 2015.

PhD in progress : Edouard Lhomme, *Analyse des déterminants de la réponse immunitaire post-vaccination dans des stratégies vaccinales expérimentales*, from Oct 2016, directed by Laura Richert.

PhD in progress : Hadrien Lorenzo, *Analyses de données longitudinales de grandes dimensions appliquées aux essais vaccinaux contre le VIH et Ebola*, from Oct 2016, co-directed by Rodolphe Thiébaud and Jérôme Saracco.

PhD in progress : Louis Capitaine, *Random forests for high dimensional longitudinal data*, from Oct 2017, co-directed by Robin Genuer and Rodolphe Thiébaud.

PhD in progress : Madelyn Rojas *Self-management of injury risk and decision support systems based on predictive computer modelling. Development, implementation and evaluation in the MAVIE cohort study*, from Oct 2017, (Injury Epidemiology team, Inserm U1219, ED SP2) co-directed by Emmanuel Lagarde, David Conesa and Marta Avalos.

PhD defense on Oct 30 2018: Chloé Pasin, *Modeling and optimizing the response to vaccines and immunotherapeutic interventions: application to Ebola virus and HIV*, from Sep 2015, co-directed by Rodolphe Thiébaud and Francois Dufour.

PhD defense on Dec 13 2018: Laura Villain "Modélisation de l'effet du traitement par injection IL7", co-directed by Daniel Commenges and Rodolphe Thiébaud, from Oct 2015.

PhD defense on Nov 16 2018: Mélanie Née *Recherche et caractérisation de profils attentionnels : mieux comprendre la place de l'attention dans la survenue des accidents de la vie courante*, from Oct 2015, co-directed by Emmanuel Lagarde, Cédric Galéra (from the research center Inserm U1219) and Marta Avalos.

10.2.3. Juries

Mélanie Prague was involved in the PhD defence jury of Steven Sanche (university of Montreal).

Mélanie Prague is a member of the follow-up dissertation comity of 3 PhD students: NICOLO Chiara, Sébastien Benzckry's PhD student (Inria Bordeaux Sud-ouest, MONC team), Marie Astrid METTEN, Jean-Francois Viel's PhD student (Universty rennes 1 Inserm U1085) and, Jonas BEAL, Sebastien Latouche's PhD student (Institut Curie Paris).

Mélanie Prague took part in the recruitment commission Inria CR Bordeaux and a postdoc recruitment committee in an European project at Pau university.

Marta Avalos is a member of the follow-up dissertation comity of 3 PhD students: Allison Singier (Pharmacoepidemiology team, Inserm U1219, ED SP2), Alexandre Conanec (Statistics, IMB, ED MI), Delphine Canzian (Education sciences, ED SP2).

Marta Avalos was involved in the PhD defence jury of Mélanie Née (University of Bordeaux).

Rodolphe Thiébaud took part in the HDR committee of Marta Avalos and the PhD defence jury of Vincent Madelain, Chloé Pasin and Laura Villain.

Daniel Commenges took part in the PhD defence jury of Laura Villain.

Laura Richert, Rodolphe Thiébaud, Robin Genuer, Boris Hejblum and Marta Avalos participated to the juries of Master in Public Health (Biostatistics, Epidemiology, Public Health)

Edouard Lhomme and Laura Richert participated to the juries of medical thesis defenses, Medical School of Bordeaux University

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- Mélanie Prague is part of the Inria Commission de Développement Technologique (CDT)
- Mélanie Prague is part of the Inria commission des emplois de recherche.
- Robin Genuer is the webmaster of the publication site of the French statistical society

10.3.2. Articles and contents

- Mélanie Prague participated to an interview for Sud ouest Eco.
- Mélanie Prague made a video of presentation of SISTM team

10.3.3. Interventions

- Chloé Pasin presented "Modélisation et optimisation de la réponse immunitaire" to L3 student from ENS Lyon visiting Inria and Hadrien Lorenzo participated to research speed meetings with these students on December 6 2018.
- Melany Durand, Hadrien Lorenzo, Chloé Pasin and Mélanie Prague participated at the "Fête de la Science" and presented "D'une goutte de sang à ta prochaine visite chez le médecin : bien personnaliser ton traitement" to high school students on October 10 2018.
- Chloé Pasin participated at the "atelier Digit'elles" with the "Femmes and Sciences" organization at the "Fête de la Science" on October 9 2018.

10.3.4. Internal action

- the whole team participated in a showcase of their activity for Inria BSO 10th anniversary on September 27th, 2018 "Une goutte de sang contient-elle plus de données qu'un smartphone ?"
- Marta Avalos and Binbin Xu conducted a workshop on "Vivre la diversité à Inria Bordeaux – Sud-Ouest", within the internal workshop for the Inria BSO 10th anniversary, June 13. Participation to the SO News of June.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [1] M. AVALOS FERNANDEZ. *Contribution to model selection via Lasso penalization in Epidemiology*, Université de Bordeaux, December 2018, Habilitation à diriger des recherches, <https://hal.archives-ouvertes.fr/tel-01964508>
- [2] C. PASIN. *Modeling and optimizing the response to vaccines and immunotherapeutic interventions. Application to Ebola virus and HIV*, Université de Bordeaux, October 2018, <https://hal.inria.fr/tel-01973077>
- [3] L. VILLAIN. *Analysing and modeling the effect of interleukin 7 in HIV-infected patients*, Université de bordeaux, December 2018, <https://hal.inria.fr/tel-01973123>

Articles in International Peer-Reviewed Journals

- [4] A. AMMASSARI, W. STÖHR, A. ANTINORI, J.-M. MOLINA, C. SCHWIMMER, P. DOMINGO, A. THALME, M. DI PIETRO, C. WALLET, A. POZNIAK, L. RICHERT, F. RAFFI. *Patient Self-Reported Adherence to Ritonavir-Boosted Darunavir Combined with Either Raltegravir or Tenofovir Disoproxil Fumarate /Emtricitabine in the Neat 001/Anrs 143 Trial*, in "Journal of Acquired Immune Deficiency Syndromes", 2018 [DOI : 10.1097/QAI.0000000000001834], <https://hal.inria.fr/hal-01888899>

- [5] I. BALELLI, V. MILISIC, G. WAINRIB. *Multi-type Galton-Watson processes with affinity-dependent selection applied to antibody affinity maturation*, in "Bulletin of Mathematical Biology", 2018, <https://arxiv.org/abs/1609.00823v2> [DOI : 10.1007/s11538-018-00548-Y], <https://hal.inria.fr/hal-01937856>
- [6] D. COMMENGES, C. ALKHASSIM, R. GOTTARDO, B. P. HEJBLUM, R. THIÉBAUT. *cytometree: A binary tree algorithm for automatic gating in cytometry analysis*, in "Cytometry Part A", 2018, bioRxiv 335554 [DOI : 10.1002/CYTO.A.23601], <https://hal.inria.fr/hal-01887966>
- [7] E. GEORGE, R. BUCCIARDINI, L. RICHERT, N. DEDES, V. FRAGOLA, P. NIEUWKERK, B. SPIRE, A. VOLNY-ANNE, B. WEST, J.-M. MOLINA, A. HORBAN, J. FOX, A. POZNIAK, S. VELLA, M. TERMOTE, F. RAFFI. *Patient-Reported Outcomes in First-Line Antiretroviral Therapy: Results from NEAT001/ANRS143 Trial Comparing Darunavir/Ritonavir in Combination with Tenofovir/Emtricitabine or Raltegravir*, in "Journal of Acquired Immune Deficiency Syndromes", 2018 [DOI : 10.1097/QAI.0000000000001844], <https://hal.inria.fr/hal-01888891>
- [8] L. GROSS, É. LHOMME, C. PASIN, L. RICHERT, R. THIÉBAUT. *Ebola vaccine development: Systematic review of pre-clinical and clinical studies, and meta-analysis of determinants of antibody response variability after vaccination*, in "International Journal of Infectious Diseases", September 2018, vol. 74, pp. 83 - 96 [DOI : 10.1016/j.ijid.2018.06.022], <https://hal.inria.fr/hal-01888913>
- [9] B. P. HEJBLUM, C. ALKHASSIM, R. GOTTARDO, F. CARON, R. THIÉBAUT. *Sequential Dirichlet Process Mixtures of Multivariate Skew t-distributions for Model-based Clustering of Flow Cytometry Data*, in "Annals Of Applied Statistics", 2018, <https://arxiv.org/abs/1702.04407> - 39 pages, 11 figures, <https://hal.inria.fr/hal-01579063>
- [10] B. P. HEJBLUM, J. CUI, L. J. LAHEY, A. CAGAN, J. A. SPARKS, J. SOKOLOVE, T. CAI, K. P. LIAO. *Association between anti-citrullinated fibrinogen antibodies and coronary artery disease in rheumatoid arthritis*, in "Arthritis Care and Research", June 2018, vol. 70, n^o 7, pp. 1113-1117 [DOI : 10.1002/ACR.23444], <https://hal.inria.fr/hal-01654403>
- [11] B. P. HEJBLUM, G. M. WEBER, K. P. LIAO, N. P. PALMER, S. CHURCHILL, N. SHADICK, P. SZOLOVITS, S. N. MURPHY, I. KOHANE, T. CAI. *Probabilistic record linkage of de-identified research datasets with discrepancies using diagnosis codes*, in "Scientific Data", 2018, <https://hal.inria.fr/hal-01942279>
- [12] S. LEFÈVRE-ARBOGAST, D. GAUDOUT, J. BENSALÉM, L. LETENNEUR, J.-F. DARTIGUES, B. P. HEJBLUM, C. FEART, C. DELCOURT, C. SAMIERI. *Pattern of polyphenol intake and the long-term risk of dementia in older persons*, in "Neurology", May 2018, vol. 90, n^o 22, pp. e1979 - e1988 [DOI : 10.1212/WNL.0000000000005607], <https://hal.inria.fr/hal-01887895>
- [13] S. LUNEMANN, A. SCHÖBEL, J. KAH, P. FITTJE, A. HÖLZEMER, A. LANGENECKERT, L. HESS, T. POCH, G. MARTRUS, W. GARCIA-BELTRAN, C. KÖRNER, A.-R. ZIEGLER, L. RICHERT, K. OLDHAFFER, J. SCHULZE ZUR WIESCH, C. SCHRAMM, M. DANDRI, E. HERKER, M. ALTFELD. *Interactions Between KIR3DS1 and HLA-F Activate Natural Killer Cells to Control HCV Replication in Cell Culture*, in "Gastroenterology", July 2018 [DOI : 10.1053/J.GASTRO.2018.07.019], <https://hal.inria.fr/hal-01888904>
- [14] C. PASIN, F. DUFOUR, L. VILLAIN, H. ZHANG, R. THIÉBAUT. *Controlling IL-7 injections in HIV-infected patients*, in "Bulletin of Mathematical Biology", August 2018, <https://hal.inria.fr/hal-01933170>

- [15] W. SALZBERGER, G. MARTRUS, K. BACHMANN, H. GOEBELS, L. HESS, M. KOCH, A. LANGE-NECKERT, S. LUNEMANN, K. OLDHAFFER, C. PFEIFER, T. POCH, L. RICHERT, C. SCHRAMM, R. WAHIB, M. BUNDERS, M. ALTFELD. *Tissue-resident NK cells differ in their expression profile of the nutrient transporters Glut1, CD98 and CD71*, in "PLoS ONE", July 2018, vol. 13, n^o 7, e0201170 p. [DOI : 10.1371/JOURNAL.PONE.0201170], <https://hal.inria.fr/hal-01888905>
- [16] J. A. SINNOTT, F. CAI, S. YU, B. P. HEJBLUM, C. HONG, I. KOHANE, K. P. LIAO. *PheProb: probabilistic phenotyping using diagnosis codes to improve power for genetic association studies*, in "Journal of the American Medical Informatics Association", May 2018 [DOI : 10.1093/JAMIA/OCY056], <https://hal.inria.fr/hal-01887930>
- [17] P. SORET, M. AVALOS FERNANDEZ, L. WITTKOP, D. COMMENGES, R. THIÉBAUT. *Lasso regularization for left-censored Gaussian outcome and high-dimensional predictors*, in "BMC Medical Research Methodology", 2018 [DOI : 10.1186/s12874-018-0609-4], <https://hal.archives-ouvertes.fr/hal-01945367>
- [18] M. TABUE-TEGUO, L. GRASSET, J. A. AVILA-FUNES, R. GENUER, C. PROUST-LIMA, K. PÉRÈS, C. FEART, H. AMIEVA, M. G.-C. HARMAND, C. HELMER, N. SALLES, M. RAINFRAY, J.-F. DARTIGUES. *Prevalence and Co-Occurrence of Geriatric Syndromes in People Aged 75 Years and Older in France: Results From the Bordeaux Three-city Study*, in "Journals of Gerontology, Series A", January 2018 [DOI : 10.1093/GERONA/GLX068], <https://hal.inria.fr/hal-01652204>
- [19] L. VILLAIN, D. COMMENGES, C. PASIN, M. PRAGUE, R. THIÉBAUT. *Adaptive protocols based on predictions from a mechanistic model of the effect of IL7 on CD4 counts*, in "Statistics in Medicine", September 2018, vol. 38, n^o 2, pp. 221-235 [DOI : 10.1002/SIM.7957], <https://hal.inria.fr/hal-01973045>
- [20] S. M. ZIEGLER, C. N. FELDMANN, S. H. HAGEN, L. RICHERT, T. BARKHAUSEN, J. GOLETZKE, V. JAZBUTYTE, G. MARTRUS, W. SALZBERGER, T. RENNÉ, K. HECHER, A. DIEMERT, P. C. ARCK, M. ALTFELD. *Innate immune responses to toll-like receptor stimulation are altered during the course of pregnancy*, in "Journal of Reproductive Immunology", August 2018, vol. 128, pp. 30 - 37 [DOI : 10.1016/J.JRI.2018.05.009], <https://hal.inria.fr/hal-01888920>

International Conferences with Proceedings

- [21] M. AVALOS FERNANDEZ, R. NOCK, C. S. ONG, J. ROUAR, K. SUN. *Representation Learning of Compositional Data*, in "NIPS 2018 - Thirty-second Conference on Neural Information Processing Systems", Montréal, Canada, S. BENGIO, H. WALLACH, H. LAROCHELLE, K. GRAUMA, N. CESA-BIANCHI, R. GARNETT (editors), Advances in Neural Information Processing Systems 31 (NIPS 2018) pre-proceedings, December 2018, vol. 31, <https://hal.archives-ouvertes.fr/hal-01945508>
- [22] M. ROJAS CASTRO, M. TRAVANCA, M. AVALOS FERNANDEZ, D. V. CONESA, L. ORRIOLS, E. LAGARDE. *MAVIE-Lab Sports: a mHealth for Injury Prevention and Risk Management in Sport*, in "DIGITAL HEALTH 2018 - 8th International Digital Health Conference", Lyon, France, April 2018, <https://hal.inria.fr/hal-01782964>

Conferences without Proceedings

- [23] H. LORENZO, J. SARACCO, R. THIÉBAUT. *Une PLS multivoie parcimonieuse avec observations manquantes pour données hétérogènes*, in "Journées de Statistique 2018 - Société Française de Statistique", EDF Lab Paris Saclay, France, May 2018, <https://hal.archives-ouvertes.fr/hal-01963431>

- [24] H. LORENZO, R. THIÉBAUT, J. SARACCO. *Multi-block high-dimensional lasso-penalized analysis with imputation of missing data applied to postgenomic data in an Ebola vaccine trial*, in "Annual workshop on Statistical Methods for Post Genomic Data - SMPGD 2018", Montpellier, France, January 2018, <https://hal.archives-ouvertes.fr/hal-01664610>
- [25] C. PASIN, L. VILLAIN, F. DUFOUR, D. COMMENGES, M. PRAGUE, R. THIÉBAUT. *Use of mathematical modeling for optimizing and adapting immunotherapy protocols in HIV-infected patients*, in "Population Approach Group Europe", Montreux, Switzerland, May 2018, <https://hal.inria.fr/hal-01934340>
- [26] M. ROJAS CASTRO, M. TRAVANCA, M. AVALOS FERNANDEZ, D. V. CONESA, E. LAGARDE. *Usefulness of Bayesian modeling in risk analysis and prevention of Home Leisure and Sport Injuries (HLIs)*, in "III Jornadas Científicas de Estudiantes de la Sociedad Española de Biometría", Bilbao, Spain, January 2018, <https://hal.archives-ouvertes.fr/hal-01964484>
- [27] M. ROJAS CASTRO, M. TRAVANCA, M. AVALOS FERNANDEZ, L. ORRIOLS, D. V. CONESA, E. LAGARDE. *Mavie-lab sports: a mhealth for injury prevention and risk management in sport*, in "The 13th World Conference on Injury Prevention and Safety Promotion (Safety 2018)", Bangkok, Thailand, November 2018, <https://hal.archives-ouvertes.fr/hal-01945586>
- [28] L. VILLAIN, D. COMMENGES, M. PRAGUE, C. PASIN, R. THIÉBAUT. *Adaptive protocols based on predictions from a mechanistic model of the effect of IL7 on CD4 counts*, in "International Biometric Conference", Baelone, Spain, July 2018, <https://hal.inria.fr/hal-01973877>

Other Publications

- [29] L. CAPITAINE, R. GENUER, R. THIÉBAUT. *Random forests for high-dimensional longitudinal data*, February 2019, <https://arxiv.org/abs/1901.11279> - working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02002979>
- [30] P. SORET, M. AVALOS FERNANDEZ, L. DELHAES, R. THIÉBAUT. *A simulation framework of high-dimensional phylogenetic microbiota data*, July 2018, 29th International Biometric Conference, Poster, <https://hal.inria.fr/hal-01856324>

References in notes

- [31] O. AALEN, K. KJETIL ROYSLAND, J. GRAN, B. LEDERGERBER. *Causality, mediation and time: a dynamic viewpoint*, in "Journal of the Royal Statistical Society: Series A (Statistics in Society)", 2007, vol. 175, n^o 4, pp. 831–861
- [32] F. CASTIGLIONE, B. PICCOLI. *Cancer immunotherapy, mathematical modeling and optimal control*, in "Biometrical Journal", 2007, vol. 247, n^o 4, pp. 723-32
- [33] R. M. GRANICH, C. F. GILKS, C. DYE, K. M. DE COCK, B. G. WILLIAMS. *Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model*, in "Lancet", 2009, vol. 373, n^o 9657, pp. 48-57, 0140 6736 English
- [34] L. HOOD, Q. TIAN. *Systems approaches to biology and disease enable translational systems medicine*, in "Genomics Proteomics Bioinformatics", 2012, vol. 10, n^o 4, pp. 181–5

- [35] Y. HUANG, D. LIU, H. WU. *Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system*, in "Biometrics", 2006, vol. 62, n^o 2, pp. 413–423
- [36] E. KUHN, M. LAVIELLE. *Maximum likelihood estimation in nonlinear mixed effects models*, in "Computational Statistics & Data Analysis", 2005, vol. 49, n^o 4, pp. 1020–1038
- [37] K.-A. LE CAO, P. MARTIN, C. ROBERT-GRANIÉ, P. BESSE. *Sparse canonical methods for biological data integration: application to a cross-platform study*, in "BMC bioinformatics", 2009, vol. 10, 34 p.
- [38] C. LEWDEN, D. SALMON, P. MORLAT, S. BEVILACQUA, E. JOUGLA, F. BONNET, L. HERIPRET, D. COSTAGLIOLA, T. MAY, G. CHÊNE. *Causes of death among human immunodeficiency virus (HIV)-infected adults in the era of potent antiretroviral therapy: emerging role of hepatitis and cancers, persistent role of AIDS*, in "International Journal of Epidemiology", 2005, vol. 34, n^o 1, pp. 121-130, 0300 5771 English
- [39] A. S. PERELSON, A. U. NEUMANN, M. MARKOWITZ, J. M. LEONARD, D. D. HO. *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time*, in "Science", 1996, vol. 271, n^o 5255, pp. 1582-6
- [40] A. S. PERELSON. *Modelling viral and immune system dynamics*, in "Nature Reviews Immunology", 2002, vol. 2, n^o 1, pp. 28-36
- [41] J. PINHEIRO, D. BATES. *Approximations to the log-likelihood function in the nonlinear mixed-effects model*, in "Journal of Computational and Graphical Statistics", 1995, vol. 4, n^o 1, pp. 12–35
- [42] B. PULENDRAN. *Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology*, in "Nature Reviews Immunology", 2009, vol. 9, n^o 10, pp. 741-7
- [43] H. PUTTER, S. HEISTERKAMP, J. LANGE, F. DE WOLF. *A Bayesian approach to parameter estimation in HIV dynamical models*, in "Statistics in Medicine", 2002, vol. 21, n^o 15, pp. 2199–2214
- [44] A. REINER, D. YEKUTIELI, Y. BENJAMINI. *Identifying differentially expressed genes using false discovery rate controlling procedures*, in "Bioinformatics", 2003, vol. 19, n^o 3, pp. 368–375
- [45] C. SCHUBERT. *Systems immunology: complexity captured*, in "Nature", 2011, vol. 473, n^o 7345, pp. 113-4
- [46] R. THIÉBAUT, B. HEJBLUM, L. RICHERT. *The analysis of "Big Data" in clinical research*, in "Epidemiology and Public Health / Revue d'Epidémiologie et de Santé Publique", January 2014, vol. 62, n^o 1, pp. 1–4 [DOI : 10.1016/J.RESPE.2013.12.021], <http://www.hal.inserm.fr/inserm-00933691>
- [47] R. THIÉBAUT, H. JACQMIN-GADDA, A. BABIKER, D. COMMENGES. *Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+cell count and HIV RNA viral load in response to treatment of HIV infection*, in "Statistics in Medicine", 2005, vol. 24, n^o 1, pp. 65-82
- [48] R. TIBSHIRANI. *Regression shrinkage and selection via the lasso*, in "Journal of the Royal Statistical Society: Series B (Statistical Methodology)", 1996, vol. 58, pp. 267–288

- [49] Y. WANG. *Derivation of various NONMEM estimation methods*, in "Journal of Pharmacokinetics and pharmacodynamics", 2007, vol. 34, n^o 5, pp. 575–593
- [50] H. WU. *Statistical methods for HIV dynamic studies in AIDS clinical trials*, in "Statistical Methods in Medical Research", 2005, vol. 14, n^o 2, pp. 171–192