RESEARCH CENTRE
Bordeaux - Sud-Ouest

IN PARTNERSHIP WITH:
INSERM, Université de Bordeaux

2020
ACTIVITY REPORT

Project-Team
SISTM

Statistics In System biology and Translational Medicine

DOMAIN
Digital Health, Biology and Earth

THEME
Modeling and Control for Life Sciences

Contents

Project-Team SISTM

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

Keywords

Computer sciences and digital sciences

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

1 Team members, visitors, external collaborators

Research Scientist

- Melanie Prague [Inria, Researcher]

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]

- Boris Hejblum [Univ de Bordeaux, Associate Professor]

- Laura Richert [Univ de Bordeaux and Univ Hospital, Professor, HDR]

Post-Doctoral Fellows

- Quentin Clairon [INSERM]

- Edouard Lhomme [Univ de Bordeaux and Univ Hospital]

- Hadrien Lorenzo [INSERM, until Aug 2020]

- Laura Villain [Inria]

PhD Students

- Marie Alexandre [Inria]

- Louis Capitaine [Univ de Bordeaux, until Dec 2020]

- Iris Beatrice Ganser [INSERM, from Sep 2020]

- Marine Gauthier [Univ de Bordeaux]

- Benjamin Hivert [INSERM, from Oct 2020]

- Helene Savel [Ipsen, CIFRE]

Technical Staff

- Houreratou Barry [INSERM, Engineer]

- Melany Durand [INSERM, Engineer]

- Fares Embouazza [INSERM, Engineer, from Nov 2020]

- Miriam Krueger [INSERM, Engineer, until Sep 2020]

- Clement Nerestan [Inria, Engineer]

- Ndoh Penn [INSERM, Engineer, from Oct 2020]

- Maria Prieto Gonzalez [INSERM, Engineer, from Sep 2020]

- Myrtille Richard [INSERM, Engineer, from Nov 2020]

- Panthea Tzourio [INSERM, Engineer, from Mar 2020]

Interns and Apprentices

- Hannah Brunner [INSERM, until Feb 2020]

- Benjamin Hivert [Inria, from Feb 2020 until Aug 2020]

- Sudip Jung Karki [INSERM, until Jun 2020]

- Ndoh Penn [INSERM, until Jul 2020]

- Myrtille Richard [INSERM, from Feb 2020 until Oct 2020]

- Arulmani Thiyagarajan [Inria, until Jun 2020]

- Helene Touchais [Inria, from Mar 2020 until Sep 2020]

Administrative Assistants

- Sandrine Darmigny [INSERM]

- Anton Ottavi [Univ de Bordeaux, Research funding coordinator]

- Audrey Plaza [Inria]

Visiting Scientists

- Gayo Diallo [Univ de Bordeaux, from Sep 2020, Inria half-delegation, HDR]

- Jane Heffernan [INSERM, from Sep 2020]

External Collaborator

- Thomas Ferte [Bordeaux Univ Hospital , from Nov 2020]

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 beneficits from a very rich ecosystem. Firstly, it is one of the rare (N=3 as of 01/2020) 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 develop and apply methods for discovering complex relationships between high dimensional data (multiblock analysis for data integration).

- To reduce data redundancy by i) high dimensional reduction ii) deconvolution.

- To visualize high dimensional data through statistically sound approaches.

- To infer cell populations abundance through gene expression data by deconvolution approach.

The first step for each type of data is to infer the maximum of information from the generated data. In fact, although data are most often high dimensional, they are not fully analyzed because of their complexity or their size. One example is the cell phenotyping where 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 unusual combination of markers. This problem is even more prevalent now with techniques that allows measurement of up to 40 markers on a single cell. We are thus extending our approach of automated gating to newly available techniques (e.g. Cytof). We need now to develop an annotation procedure and we want to liaise with researches on ontologies.

However, measuring specific cells with many markers required a lot of blood, preferentially fresh and therefore it is very difficult to implement the measures to large sample size with many repeated measurements. This is why, we want to look at how much cell phenotyping could be replaced 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: to improve 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, the analysis requests new approaches to take into account several methodological difficulties such as heavy 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).

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. Therefore, we would like to explore new approaches such as deep Collective Matrix Factorization (dCMF). It is an unsupervised deep learning model that can seamlessly learn latent factors from multiple heterogeneous relational data sources and use them to complete each matrix in the collection (Mariappan

and Rajan, 2019). We want to extend this approach to repeated measures in collaboration with the author (Vaibhav Rajan, Singapor).

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 [11].

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 (Coliin 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), 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) 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 plateforms 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" [61], "Systems vaccinology" [60], "Systems medicine" [58]. 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 [59]. 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) [57]. 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. 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.

5 Highlights of the year

5.1 Ebola vaccine development

Beyond the scientific publications including the results of EBL2001 trial (Lancet Infectious Diseases), the Ad26.ZEBOV/MVA-BN-Filo developped with Janssen has been approved by FDA and EMA for emergency use. The submitted file included the model predictions of the duration of the vaccine response.

5.2   Response to COVID19

The team is very much involved in the response to COVID19 pandemics. By chronological orders, here are the main projects that have been started:

1) Modelling of the COVID19 epidemics in France (GESTEPID) including the development of an optimal control approach based on reinforcement learning. Preprint:

https://www.medrxiv.org/content/10.1101/2020.04.21.20073536v1

2) Analyzing the early immune response in hospitalized patients from the french COVID19 cohort. This has been done in the context of the WP5 Systems biology of the new consortium Corona Accelerated R&D in Europe (CARE). The team is also involved in the WP8 data management through the settlement of the Labkey based data warehouse and the WP7 on clinical trials. Preprint:

https://www.medrxiv.org/content/10.1101/2020.12.12.20246934v1

3) In the context of the Vaccine Research Institute, the team is contributing to the development of a new COVID19 vaccine based on targeting dendritic cells. We have estimated the effect of the vaccine from initial NHP data. Preprint:

https://www.researchgate.net/publication/349460203_Targeting_SARS-CoV-2_recep
tor-binding_domain_to_cells_expressing_CD40_improves_protection_to_infection_in_c
onvalescent_macaques.

6   New software and platforms

6.1   New software

6.1.1   dearseq

Keywords: Biostatistics, Bioinformatics, Computational biology, Statistics

Functional Description: Package R for analyzing RNA-seq data. The 2 main functions of the package are dear_seq() and dgsa_seq(): Gene-wise Differential Analysis of RNA-seq data can be performed using the function dear_seq(). Gene Set Analysis of RNA-seq data can be performed using the function dgsa_seq().

URL: https://github.com/borishejblum/dearseq

Authors: Boris Hejblum, Marine Gauthier

Contact: Boris Hejblum

Partner: RAND Corporation

6.1.2   VICI

Name: Vaccine Induced Cellular Immunogenicity with Bivariate Modeling

Keywords: Biostatistics, Data analysis

Functional Description: Package R - Vaccine Induced Cellular Immunogenicity with Bivariate Modeling https://shiny-vici.apps.math.cnrs.fr/

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

Authors: Boris Hejblum, Clement Nerestan

Contacts: Boris Hejblum, Edouard Lhomme, Clement Nerestan

6.1.3   CBCtool

Name: R package CBCtool

Keywords: Biostatistics, Biological sequences, Machine learning, Discrimination, Sparse regularization, Regression, Polynomial regression

Functional Description: We propose a cost-sensitive Lasso-penalized additive logistic regression to identify which CBC variables are associated with a higher risk of abnormal manual smear and at which cutoff values. Additive functions are considered to belong to the space of piecewise constant functions.

URL: https://github.com/mavalosf/CBCtool

Contacts: Marta Avalos Fernandez, Hélène Touchais

6.1.4   ccdf

Keyword: Biostatistics

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

URL: https://github.com/Mgauth/ccdf

Contacts: Boris Hejblum, Marine Gauthier

6.1.5   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

Contacts: Cedric Colas, Clément Moulin-Frier, Melanie Prague

7   New results

The year 2020 was marked by the covid crisis and its impact on society and its overall activity. The world of research was also greatly affected: Faculty members have seen their teaching load increase significantly; PhD students and post-docs have often had to deal with a worsening of their working conditions, as well as with reduced interactions with their supervisors and colleagues; most scientific collaborations have been greatly affected, with many international and national activities cancelled or postponed.

On the other hand, several researchers of the team have been heavily mobilised for Covid research projects from March 2020 onwards.

7.1 High-dimensional and statistical learning

7.1.1 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 has been extended (Gautier, NAR 2020 [22]). 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. Interestingly, our latest results have shown that even simple gene by gene differential analyses were subject to a huge type-1 error and we therefore have proposed a new statistical approach which is controlling type 1 and type 2 error much more efficiently that the three most popular approaches used in the world (EdgeR, DESeq2, and limma-voom). Those methodologies have been applied to deepen our understanding of SARS-CoV-2 infection (Bouadma, Journal of Clinical Immunlogy, 2020 [13]).

7.1.2 Random Forests for High-Dimensional Data

The integrative work on random forests, a nowadays very popular machine learning approach, for which we have developed a variable selection approach (Genuer, R Journal 2015), extended to big data (Genuer, Bid Data Research 2017), extended to longitudinal data (Capitaine, SMMR, 2020 [15]). Further, a book for R users has been published (R. Genuer, J.-M. Poggi, Random Forests with R, Springer Nature eds [41]). The work done on this method is important because the performances of the random forest make their usefulness and dissemination huge.

7.1.3 Lasso-type penalties for High-Dimensional Data

M. Avalos and H. Touchais (intern in the axis "High-dimensional and statistical learning", supervised by M. Avalos) developed a machine learning approach to optimise criteria for manual smear review following automated blood count analysis, in collaboration with M. Henriquez (ELSA Clinical Laboratory, IntegraMedica, part of BUPA and Cardio MR, Millennium Science Initiative Program, ANID, Santiago, Chile). The poster: "A decision-making tool to fine-tune abnormal levels in the complete blood count tests" was presented in the ML4H - Machine Learning for Health workshop at NeurIPS 2020 [38]. The associated R package CBCtools, allowing to identify which variables from the complete blood count (CBC) are associated with a higher risk of abnormal manual smear and at which cutoff values, is publicly available

7.2 Mechanistic learning

7.2.1 Ebola models

New models have been developed for the response to the Ebola vaccine. In aparticular, a new model including the B cell memory response has been defined and its mathematical proprieties have been published. We now focus on its estimation from phase 2 clinical data using EBL2001 and EBL1001. Meanwhile, we evaluate the predictive abilities of the model published by Pasin et al. (2019) on long-term follow-up data up to more than 2 years using EBL3001. Two articles are under writing.

New publication: Balelli et al. (2020). A model for establishment, maintenance and reactivation of the immune response after vaccination against Ebola virus. Journal of theoretical biology [11].

From the very begining of the SARS-COV-2/COVID19 pandemics, we got involved in modeling the epidemics and the effect of non-pharmaceutical measures. Our work was cited in the report from the scientific comittee on June 2nd. Moreover, we also applied our within host models for viral and antibodies dynamics to available data and more recently data from a new vaccine platform in France within the vaccine research institute. Several papers are under writing.

7.2.2 Estimation method

A new approach to estimate model parameters using a regularization method based on the control theory has been submitted for publication. It mitigates the effect of model misspecification on estimation accuracy and avoid estimation of initial conditions. On the other hand, we published two comparison

papers: one comparing mechanistic estimations and splines approaches to model HIV viral dynamics and one comparing mechanitsic approaches and learning methods to predict tumor growth in breast cancer patients. New publication: Bing et al. (2020). Comparison of empirical and dynamic models for HIV viral load rebound after treatment interruption. Statistical Communications in Infectious Diseases [12]. Nicolò et al (2020). Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer. JCO clinical cancer informatics [29].

7.3 Translational vaccinology

7.3.1 Vaccine clinical trials

Three papers with the primary results of vaccine trials have been published in 2020 :
    a) Ebola vaccine trial (Pollard et al, Lancet Inf Dis 2020) [32];
    b) Malaria vaccine trial (Sirima et al, Lancet Inf Dis 2020) [32] ; and
    c) Pertussis vaccine trial (Jahnmatz et al, Lancet Inf Dis 2020).
    In addition, the randomized Ebola vaccine trial Prevac, that assesses the immunogenicity of three different Ebola vaccine strategies in four African countries, has reached key milestones in 2020, with preliminary month 12 results becoming available. An immunological sub-study that is nested in the trial in Guinea and is conducted by the VRI and analyzed by the SISTM team has shown the response profiles of cytokine-producing T-cells after vaccination as well as plasma cytokine responses. Further data from this sub-study, in particular gene expression data, will be analyzed in 2021.
    A new first-in-human phase I vaccine trial of a novel HIV vaccine concept has also been designed and set-up with the VRI in 2020 (ANRS VRI06 trial). Trial enrolment will start in 2021.

7.3.2 Covid treatment trials

The researchers of the team have been heavily mobilised for Covid clinical research projects from March 2020 onwards. Particular high-profile projects are a) the Coverage clinical trial; and b) the IMI-2 Care consortium.
    Coverage is a 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 is still ongoing in 2021 and has been adapted several times since its start. The Sistm team has coordinated the Coverage-Immuno study (financed by EIT Health), nested in the main trial. The Coverage-Immuno study enrolled 23 participants in 2020 and used an innovative sampling methods allowing outpatients to perform very regular blood sampling for RNA Seq analyses at home themselves. Cell population and cytokine data have already been analysed by the team, and the additional RNA Seq data (available in early 2021) will provide an excellent opportunity to gain insights in the pathophysiology at early Covid disease stages.

7.3.3 Systems immunology analyses

We have also advanced the integrative statistical analyses of the immunogenicity data generated in the IMI-2 Ebovac2 consortium (Ebola vaccine trials), for which we were the designated task leader. The analyses, using random forests, have allowed to assess the effects of the vaccine on gene expression, as well as to identify early gene expression (at day 1 after the first vaccination) as early predictors of later antibody responses (21 days after a 2nd vaccine injection one to three months later). Results have been presented to the international consortium, and a manuscript will be prepared for publication.

7.3.4 Methodological developments for vaccine trials

The statistical methods developed by the team for the analysis of functional T-cell data using a bivariate modelling approach has been published in the Journal of Immunological Methods in 2021. 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.

8 Partnerships and cooperations

8.1 International initiatives

8.1.1 Inria International Labs

DYNAMHIC

Title: *Dynamical modeling of HIV Cure*

Duration: 2019 - 2023

Coordinator: Melanie Prague

Partners:
> Harvard Program for Evolutionary Dynamics, Harvard University (United States)

Inria contact: Melanie Prague

Summary: The aim of the DYNAMHIC Associate Team is to bring together a mathematical biology team at Harvard and the Inria team SISTM of applied statisticians at Bordeaux Sud-ouest. This collaboration will allow the analysis of unique pre-clinical non human primates data of HIV cure interventions. In particular, we will focus on immunotherapy and therapeutic vaccine, which are very promising in term of efficacy and are at the leading edge of pre-clinical research in the area. The novelty of the approach is to propose an integrative project studying complex biological processes with novel mathematical statistical models, which has the potential to yield predictive computational tools to assist in the design of both therapeutic products and clinical trials for HIV cure. Finally, the associate team is the opportunity to provide the research group with an official administrative framework. And, to continue to develop a promising research topic connected but different from those funded up to now.

SWAGR

Title: *Statistical Workforce for Advanced Genomics using RNAseq*

Duration: 2018 - 2020

Coordinator: Boris Hejblum

Partners:
> Statistics Group, RAND Corporation (United States)

Inria contact: Boris Hejblum

Summary: 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.

8.1.2 Participation in other international programs

NIPAH (Chine) – Programme de coopération scientifique France/Chine. (2019-2022) - M. Prague is workpackage co-PI - Sino-French Agreement Aviesan. Sept. 2018 – Aug. 2023, 150,000€. 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.

amfAR (USA) – "Mechanistic and empirical modeling of viral rebound to identify predictors of post-treatment control". Contractor M. Prague, Jan. 2019 – Jan 2020, 4,000€ Travel Grant.

8.2 International research visitors

8.2.1 Visits of international scientists

Jane Heffernan, professor of mathematical modelling at York University of Toronto is an invited professor (from Sept 2020 to Aug 2021).

8.2.2 Visits of national scientists

Gayo Diallo from Univ de Bordeaux, Inria half-delegation, from Sep 2020

8.3 European initiatives

8.3.1 FP7 & 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), Universitaet-sklinikum 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.

EHVA (`https://www.ehv-a.eu`): European HIV Vaccine Alliance: a EU platform for the discovery and evaluation of novel prophylactic and therapeutic vaccine candidates. Coordinator: Inserm/University of Lausanne. Other partners: EHVA consortium gathers 41 partners. R. Thiébaut. Duration: 60 months. 01/01/2016 - 31/12/20 – 208 686 euros.

8.3.2 Collaborations in European programs, except FP7 and H2020

EBOVAC2 (`https://www.ebovac2.com`): 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 & 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 € + EFPIA 50 710 893 €. Amount for SISTM: 2 930 196 euros

COVERAGE-Immuno : funded by EIT-Health, it is coordinated by the Institut National de la Santé et de la Recherche Médicale (Inserm) in collaboration with the Institut National de Recherche en Informatique et en Automatique (Inria) which aims at performing a deep, repeated evaluation of immunological markers and transcriptomics data in Covid-19+ patients treated at home in the context of COVERAGE, a randomized clinical trial evaluating several experimental treatments at home (legal trial sponsor:

Bordeaux University Hospital). Duration; 9 months. 18/04/2020 – 31/12/2020. Total amount: EIT 471 138 € including 88 638 Euros for SISTM.

EBOVAC1: Development of a Prophylactic Ebola Vaccine Using an Heterologous Prime-Boost Regimen. Coordinated by London School of Hygiene & 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 & 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&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 heath 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&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 turn-around time. In this project, SISTM will be involved in statistical analysis of the tests data and will lead the WP3 which aims at modelling 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) http://vaccine-research-institute.fr.

- 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.

- SIDACTION: Towards HIV functional cure, down selection of immunotherapeutic strategies using an HIV/HIS mice model (2019-2021) (R. Thiébaut) 18 000 euros.

- Ecole Universitaire de Recherche « Digital Public Health » PIA3 –Bordeaux -University - 2018-2028 – Head : R. Thiébaut – budget : 4 517 700 euros.

8.4.1 Expert Appraisals

- Rodolphe Thiébaut is a member of the CNU 46.04 (Biostatistiques, informatique médicale et technologies de communication).

- Rodolphe Thiébaut 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 CSS13 (Recherches cliniques et physiopathologiques dans l'infection à VIH).

- Laura Richert is an expert for the PHRC (Programme hospitalier de recherche Clinique) and was an expert on the Flash Covid research call by DGOS in 2020.

- Marta Avalos is an expert for the ANSM (Agence nationale de sécurité du médicament et des produits de santé)

8.4.2 Various Partnership

The project team members are involved in:

- 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).

- F-CRIN (French clinical research infrastructure network), initiated in 2012 by ANR under "Programme des Investissements d'avenir". (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).

- Collaboration with Inserm PRC (pôle Recherche clinique).

- Collaboration with Inserm Reacting (REsearch and ACTion targeting emerging infectious diseases) network

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

9 Dissemination

9.1 Promoting scientific activities

- 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).

- Hélène Savel is a board member of the "Biopharmacie et Santé" group of SFdS.

- Marta Avalos is general secretary of the "Statistics and Sport" group of SFdS.

- Mélanie Prague is co-president of the communication group of SFdS, in charge of redefining de condition of sponsoring SfdS by enterprises.

9.1.1 Scientific events: organisation

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

- Rodolphe Thiébaut was a member of the scientific committee of the national conference on clinical research (EPICLIN) from 2017 and Laura Richert from 2020

- Rodolphe Thiébaut is a member of the scientific committee of the IWHOD International Workshop on HIV Observational Databases since 2013 (`http://iwhod.org/Committee`) and chair in 2020-2021.

9.1.2 Scientific events: selection

Member of the conference program committees

- Marta Avalos was a member of the Program Committee of the Conférence d'Apprentissage–CAp20, ACM Conference on Health, Inference, and Learning, 2020, Workshop Machine Learning for Health–ML4H at NeurIPS.

- Boris Hejblum was a member of Scientific Programme Committee of the 42nd Conference of the International Society for Clinical Biostatistics (ISCB).

Reviewer

- Marta Avalos was a reviewer for NeurIPS 2020.

9.1.3 Journal

Member of the editorial boards

- Melanie Prague is associate editor of "International journal of Biostatistics" (since 2018)

- Rodolpe Thiébaut is section editor of IMIA Yearbook in Medical Informatics (since 2017)

Reviewer - reviewing activities

- AIDS (Rodolphe Thiébaut)

- Annals of Applied Statistics (Boris Hejblum)

- Am J Public Health (Mélanie Prague)

- Biometrics (Mélanie Prague, Boris Hejblum)

- Bayesian Analysis (Boris Hejblum)

- Bioinformatics (Boris Hejblum)

- Computational Statistics and Data Analysis (Boris Hejblum)

- IMIA Yearb Med Inform (Marta Avalos)

- Journal of Computation Statistics and Data Analysis (Boris Hejblum)

- Journal of Computational and Graphical Statistics (Marta Avalos)

- Journal of the American Statistical Association (Robin Genuer)

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

- Journal of the Royal Statistical Society B (Mélanie Prague)

- Operations research (Robin Genuer)

- PLOS Computational Biology (Boris Hejblum)

- Society of clinical trial (Mélanie Prague)

- STAT (Boris Hejblum)

- Statistical Methods in Medical Research (Mélanie Prague)

- Statistical science (Mélanie Prague)

- Trials (Laura Richert)

9.1.4   Invited talks

"Apprentissage non-supervisé pour le traitement de données de cytométrie en flux" at *Journées 2020 conjointes Groupe De Recherche "Statistiques & Santé" — Société Française de Biométrie — groupe Biopharma de la Société Française de Statistique* (Boris Hejblum).

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 CNU 46.04 (Biostatistiques, informatique médicale et technologies de communication).
    Rodolphe Thiébaut is a member of the Scientific Council of Inserm.
    Rodolphe Thiébaut is a member of the commitee "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 and Sud Sida-HIV Hépatites) in the CCS13 (Recherches cliniques et physiopathologiques dans l'infection à VIH).
    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).

9.1.6 Research administration

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).

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 teach in the Master of Public Health (M1 Santé publique, M2 Biostatistique and/or M2 Epidemiology and/or M2 Public Health Data Science) and the Digital Public Health graduate program, University of Bordeaux.
  - Master: Marta Avalos, Robin Genuer, Louis Capitaine and Marine Gautier teach 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 and Edouard Lhomme teach in PACES and DFASM1-3 for Medical degree at Univ. Bordeaux.
  - Master: Laura Richert and Edouard Lhomme teach in the Master of Vaccinology from basic immunology to social sciences of health (University Paris-Est Créteil, UPEC).
  - Teaching unit coordination: Laura Richert, 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.

- **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". Several PhD students in the team teach in this ODL. Mélanie Prague teaches in the Diplôme universitaire "Méthodes statistiques de régression en épidémiologie".
  - ODL University Course: Laura Richert co-coordinates and teaches in the Diplôme universitaire "Recherche Clinique".
  - ODL University project: Robin Genuer participated to the IdEx Bordeaux University "Défi numérique" project "BeginR" (`http://beginr.u-bordeaux.fr`).

9.2.2 Supervision

- PhD: Louis Capitaine, *Random forests for high-dimensional longitudinal data*, defence date 17th Dec, directed by Robin Genuer and Rodolphe Thiébaut.

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

- 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: 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 Sept 2020.

- 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 (Inserm) and Marta Avalos.

- Several team's members (co-)supervised Master 2 (Hannah Brunner, Benjamin Hivert, Sudip Jung Karki, Ndoh Penn, Myrtille Richard, Arulmani Thiyagarajan, Hélène Touchais) and Master 1 (Alexis Francois, Lucie Ferrand) internship students.

9.2.3 Juries

- Marta Avalos was involved in the PhD defence juries of Nadim Ballout (University Claude Bernard Lyon 1) and Johann Faouzi (Sorbonne University and Paris brain institut).

- Laura Richert was involved in the PhD defence jury of Elena Gonçalves (Sorbonne University, Paris).

- Robin Genuer and Rodolphe Thiébaut were involved in the PhD defence jury of Louis Capitaine (University of Bordeaux) as PhD supervisors.

- Mélanie Prague was involved in the PhD defence jury of Johann Faouzi (Sorbonne University and Paris brain institut).

- Marta Avalos is a member of the follow-up dissertation committee of 1 PhD student: Alexandre Conanec (Statistics, IMB, ED MI).

- Mélanie Prague is a member of the follow-up dissertation committee of 1 PhD student: Imke Mayer (Ecole polytechnique).

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

- Marta Avalos participated to the juries of Master of Applied Mathematics and Statistics (2nd year), University of Bordeaux.

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

9.3   Popularization

Rodolphe Thiébaut participated to Interstices [55].

Rodolphe Thiébaut has been interviewed on Covid-19: Le Parisien 10/04/2020, La Croix 5/5/2020, Sud Ouest 25/5/2020, Le Parisien 08/06/2020, University of Bordeaux: https://www.u-bordeaux.fr/Actualites/De-la-recherche/Sante-publique-une-recherche-pluridisciplinaire-et-transversale and participated to the Digital Aquitaine Webinaire "Covid-19, la donnée au service de la gestion de la crise" https://digital-aquitaine.com/calendar/webinaire-covid-19-la-donnee-au-service-de-la-gestion-de-la-crise/. Rodolphe Thiébaut co-organized a conference CNRS INS2I "Intelligence artificielle et santé", Paris (23/01/2020)

Boris Hejblum presented "Apprentissage non supervisé dans le domaine du biomédical" at the AI4industry 2020 workshop.

Marta Avalos participated to the "3e Journée Dataquitaine : IA, RO et Data Science", Talence [56].

9.3.1   Internal or external Inria responsibilities

Mélanie prague is part of two committees at Inria: "ADT - Aide au développement technologique" and "CER - Comission Emploi recherche".

10   Scientific production

10.1   Major publications

[1]   D. Agniel and B. P. Hejblum. 'Variance component score test for time-course gene set analysis of longitudinal RNA-seq data'. In: *Biostatistics* 18.4 (2017), pp. 589–604. URL: https://hal.inria.fr/hal-01579077.

[2]   D. Commenges, C. Alkhassim, R. Gottardo, B. P. Hejblum and R. Thiébaut. 'cytometree: A binary tree algorithm for automatic gating in cytometry analysis'. In: *Cytometry Part A* 93.11 (2018), pp. 1132–1140. URL: https://hal.inria.fr/hal-01887966.

[3]   D. Commenges and H. Jacqmin-Gadda. *Dynamical Biostatistical Models*. Chapman and Hall/CRC, 2015. URL: https://hal.inria.fr/hal-01580149.

[4]   A. Jarne, D. Commenges, M. Prague, Y. Levy and R. Thiébaut. 'Modeling CD4 + T cells dynamics in HIV-infected patients receiving repeated cycles of exogenous Interleukin 7'. In: *Annals of Applied Statistics* (2017). URL: https://hal.inria.fr/hal-01579008.

[5]   C. Pasin, F. Dufour, L. Villain, H. Zhang and R. Thiébaut. 'Controlling IL-7 injections in HIV-infected patients'. In: *Bulletin of Mathematical Biology* (2018).

[6]   M. Prague, D. Commenges, J. M. Gran, B. Ledergerber, J. Young, H. Furrer and R. Thiébaut. 'Dynamic models for estimating the effect of HAART on CD4 in observational studies: Application to the Aquitaine Cohort and the Swiss HIV Cohort Study'. In: *Biometrics* (2017). URL: https://hal.inria.fr/hal-01406614.

[7]   A. Rechtien, L. Richert, H. Lorenzo, G. Martrus, B. P. Hejblum, C. Dahlke, R. Kasonta, M. Zinser, H. Stubbe, U. Matschl, A. Lohse, V. Krähling, M. Eickmann, S. Becker, R. Thiébaut, M. Altfeld and M. Addo. 'Systems Vaccinology Identifies an Early Innate Immune Signature as a Correlate of Antibody Responses to the Ebola Vaccine rVSV-ZEBOV'. In: *Cell Reports* 20.9 (2017), pp. 2251–2261. URL: https://hal.inria.fr/hal-01579246.

[8]   L. Villain, D. Commenges, C. Pasin, M. Prague and R. Thiébaut. 'Adaptive protocols based on predictions from a mechanistic model of the effect of IL7 on CD4 counts'. In: *Statistics in Medicine* 38.2 (2018), pp. 221–235.

10.2 Publications of the year

International journals

[9] S. Ajana, A. Cougnard-Grégoire, J. Colijn, B. M. Merle, T. Verzijden, P. T. M. de Jong, A. Hofman, J. Vingerling, B. Hejblum, J.-F. Korobelnik, M. Meester-Smoor, H. Jacqmin-Gadda, C. C. Klaver, C. Delcourt and . EYE-RISK consortium. 'Predicting Progression to Advanced Age-Related Macular Degeneration from Clinical, Genetic, and Lifestyle Factors Using Machine Learning'. In: *Ophthalmology: Journal of The American Academy of Ophthalmology* (2020). DOI: 10.1016/j.ophtha.2020.08.031. URL: https://hal.inria.fr/hal-03100320.

[10] P. Avery, Q. Clairon, R. Henderson, C. James Taylor and E. Wilson. 'Robust and adaptive anticoagulant control'. In: *Journal of the Royal Statistical Society: Series C Applied Statistics* 69.3 (June 2020), pp. 503–524. DOI: 10.1111/rssc.12403. URL: https://hal.archives-ouvertes.fr/hal-03152242.

[11] I. Balelli, C. Pasin, M. Prague, F. Crauste, T. Van Effelterre, V. Bockstal, L. Solforosi and R. Thiébaut. 'A model for establishment, maintenance and reactivation of the immune response after vaccination against Ebola virus'. In: *Journal of Theoretical Biology* 495 (2020), p. 110254. DOI: 10.1016/j.jtbi.2020.110254. URL: https://hal.archives-ouvertes.fr/hal-03160892.

[12] A. Bing, Y. Hu, M. Prague, A. Hill, J. Li, R. Bosch, V. DeGruttola and R. Wang. 'Comparison of empirical and dynamic models for HIV viral load rebound after treatment interruption'. In: *Statistical Communications in Infectious Diseases* (17th Aug. 2020). DOI: 10.1515/scid-2019-0021. URL: https://hal.archives-ouvertes.fr/hal-02942247.

[13] L. Bouadma, A. Wiedemann, J. Patrier, M. Surénaud, P.-H. Wicky, E. Foucat, J.-L. Diehl, B. Hejblum, F. Sinnah, E. de Montmollin, C. Lacabaratz, R. Thiébaut, J. F. Timsit and Y. Lévy. 'Immune Alterations in a Patient with SARS-CoV-2-Related Acute Respiratory Distress Syndrome'. In: *Journal of Clinical Immunology* 40 (2020), pp. 1082–1092. DOI: 10.1007/s10875-020-00839-x. URL: https://hal.inria.fr/hal-03160737.

[14] S. Bouasker, W. Inoubli, S. Ben Yahia and G. Diallo. 'Pregnancy Associated Breast Cancer gene expressions : new insights on their regulation based on Rare Correlated Patterns'. In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2020). DOI: 10.1109/TCBB.2020.3015236. URL: https://hal.archives-ouvertes.fr/hal-03145240.

[15] L. Capitaine, R. Genuer and R. Thiébaut. 'Random forests for high-dimensional longitudinal data'. In: *Statistical Methods in Medical Research* (2020). DOI: 10.1177/0962280220946080. URL: https://hal.archives-ouvertes.fr/hal-03066138.

[16] M. Chalouni, J. Rodriguez-Centeno, A. Samri, J. Blanco, N. Stella-Ascariz, C. Wallet, H. Knobel, D. Zucman, B. Alejos Ferreras, R. Thiébaut, B. Autran, F. Raffi and J. R. Arribas. 'Correlation between blood telomere length and CD4+ CD8+ T-cell subsets changes 96 weeks after initiation of antiretroviral therapy in HIV-1-positive individuals: PLoS One'. In: *PLoS ONE* 15.4 (2020), 9 p. URL: https://hal.archives-ouvertes.fr/hal-03160741.

[17] Q. Clairon, R. Henderson, N. Young, E. Wilson and C. James Taylor. 'Adaptive treatment and robust control'. In: *Biometrics* (2020). DOI: 10.1111/biom.13268. URL: https://hal.archives-ouvertes.fr/hal-03152210.

[18] M. Coppry, C. Leroyer, M. Saly, A. G. Venier, C. Slekovec, X. Bertrand, S. Parer, S. Alfandari, E. Cambau, B. Mégarbane, C. Lawrence, B. Clair, A. Lepape, P. Cassier, D. Trivier, A. Boyer, H. Boulestreau, J. Asselineau, V. Dubois, R. Thiébaut and A. M. Rogues. 'Exogenous acquisition of Pseudomonas aeruginosa in intensive care units: a prospective multi-centre study (DYNAPYO study)'. In: *Journal of Hospital Infection* 104.1 (Jan. 2020), pp. 40–45. DOI: 10.1016/j.jhin.2019.08.008. URL: https://hal.archives-ouvertes.fr/hal-02541488.

[19] M. Coppry, C. Leroyer, M. Saly, A. G. Venier, C. Slekovec, X. Bertrand, S. Parer, S. Alfandari, E. Cambau, B. Mégarbane, C. Lawrence, B. Clair, A. Lepape, P. Cassier, D. Trivier, A. Boyer, H. Boulestreau, J. Asselineau, V. Dubois, R. Thiébaut and A.-M. Rogues. 'Exogenous acquisition of Pseudomonas aeruginosa in intensive care units: a prospective multicentre study, DYNAPYO study'. In: *The Journal of hospital infection* 104.1 (2020), pp. 40–45. DOI: 10.1016/j.jhin.2019.08.008. URL: https://hal.archives-ouvertes.fr/hal-03161089.

[20] S. Cossin and R. Thiébaut. 'Public Health and Epidemiology Informatics: Recent Research Trends Moving toward Public Health Data Science: Yearb Med Inform'. In: *IMIA Yearbook of Medical Informatics* 29.1 (2020), pp. 231–234. DOI: 10.1055/s-0040-1702020. URL: https://hal.archives-ouvertes.fr/hal-03116855.

[21] T. Ferté, S. Cossin, T. Schaeverbeke, T. Barnetche, V. Jouhet and B. Hejblum. 'Automatic phenotyping of electronical health record: PheVis algorithm'. In: *Journal of Biomedical Informatics* (2021). DOI: 10.1016/j.jbi.2021.103746. URL: https://hal.inria.fr/hal-03100435.

[22] M. Gauthier, D. Agniel, R. Thiébaut and B. Hejblum. 'dearseq: a variance component score test for RNA-Seq differential analysis that effectively controls the false discovery rate'. In: *NAR Genomics and Bioinformatics* 2.4 (2020), lqaa093. DOI: 10.1093/nargab/lqaa093. URL: https://hal.inria.fr/hal-02138664.

[23] S. H. Hagen, F. Henseling, J. Hennesen, H. Savel, S. Delahaye, L. Richert, S. M. Ziegler and M. Altfeld. 'Heterogeneous Escape from X Chromosome Inactivation Results in Sex Differences in Type I IFN Responses at the Single Human pDC Level'. In: *Cell Reports* 33.10 (Dec. 2020), p. 108485. DOI: 10.1016/j.celrep.2020.108485. URL: https://hal.inria.fr/hal-03151773.

[24] L. Hess, G. Martrus, A. Ziegler, A. Langeneckert, W. Salzberger, H. Goebels, A. Sagebiel, S. Hagen, T. Poch, G. Ravichandran, M. Koch, C. Schramm, K. Oldhafer, L. Fischer, G. Tiegs, L. Richert, M. Bunders, S. Lunemann and M. Altfeld. 'The Transcription Factor Promyelocytic Leukemia Zinc Finger Protein Is Associated With Expression of Liver-Homing Receptors on Human Blood CD56 bright Natural Killer Cells'. In: *International Hepatology Communications* 4.3 (Mar. 2020), pp. 409–424. DOI: 10.1002/hep4.1463. URL: https://hal.inria.fr/hal-02941498.

[25] M. Iannetta, S. Isnard, J. Manuzak, J.-B. Guillerme, K. Bailly, M. Notin, M. Andrieu, S. Amraoui, L. Vimeux, S. Figueiredo, B. Charmeteau-De Muylder, L. Vaton, A. Samri, E. X. Hatton, B. Autran, R. Thiébaut, N. Chaghil-Boissière, D. Glohi, C. Charpentier, D. Descamps, F. Brun-Vézinet, S. Matheron, R. Cheynier and A. Hosmalin. 'Conventional Dendritic Cells and Slan(+) Monocytes During HIV-2 Infection'. In: *Frontiers in Immunology* 11 (2020), p. 1658. DOI: 10.3389/fimmu.2020.01658. URL: https://hal.archives-ouvertes.fr/hal-03177503.

[26] Y. Lévy, C. Lacabaratz, K. Ellefsen-Lavoie, W. Stöhr, J.-D. Lelièvre, P.-A. Bart, O. Launay, J. Weber, B. Salzberger, A. Wiedemann, M. Surénaud, D. Koelle, H. Wolf, R. Wagner, V. Rieux, D. Montefiori, N. Yates, G. Tomaras, R. Gottardo, B. Mayer, S. Ding, R. Thiébaut, S. McCormack, G. Chêne and G. Pantaleo. 'Optimal priming of poxvirus vector (NYVAC)-based HIV vaccine regimens for T cell responses requires three DNA injections. Results of the randomized multicentre EV03/ANRS VAC20 Phase I/II Trial'. In: *PLoS Pathogens* 16.6 (26th June 2020), e1008522. DOI: 10.1371/journal.ppat.1008522. URL: https://hal.inria.fr/hal-03142733.

[27] É. Lhomme, B. Hejblum, C. Lacabaratz, A. Wiedemann, J.-D. Lelièvre, Y. Lévy, R. Thiébaut and L. Richert. 'Analyzing cellular immunogenicity in vaccine clinical trials: a new statistical method including non-specific responses for accurate estimation of vaccine effect'. In: *Journal of Immunological Methods* 477 (2020), p. 112711. DOI: 10.1016/j.jim.2019.112711. URL: https://hal.inria.fr/hal-03161945.

[28] L. Lin and B. Hejblum. 'Bayesian mixture models for cytometry data analysis'. In: *Wiley Interdisciplinary Reviews: Computational Statistics* e1535 (16th Oct. 2020). DOI: 10.1002/wics.1535. URL: https://hal.inria.fr/hal-03100447.

[29] C. Nicolò, C. Périer, M. Prague, C. Bellera, G. Macgrogan, O. Saut and S. Benzekry. 'Machine Learning and Mechanistic Modeling for Prediction of Metastatic Relapse in Early-Stage Breast Cancer'. In: *JCO Clinical Cancer Informatics* 4 (Sept. 2020), pp. 259–274. DOI: 10.1200/CCI.19.00133. URL: https://hal.archives-ouvertes.fr/hal-02942177.

[30] C. Pflitsch, C. N. Feldmann, L. Richert, S. Hagen, A. Diemert, J. Goletzke, K. Hecher, V. Jazbutyte, T. Renné, P. C. Arck, M. Altfeld and S. Ziegler. 'In-depth characterization of monocyte subsets during the course of healthy pregnancy'. In: *Journal of Reproductive Immunology* 141 (Sept. 2020), p. 103151. DOI: 10.1016/j.jri.2020.103151. URL: https://hal.inria.fr/hal-02941491.

[31] V. Schwane, V. H. Huynh-Tran, S. Vollmers, V. M. Yakup, J. Sauter, A. Schmidt, S. Peine, M. Altfeld, L. Richert and C. Körner. 'Distinct Signatures in the Receptor Repertoire Discriminate CD56bright and CD56dim Natural Killer Cells'. In: *Frontiers in Immunology* 11 (1st Dec. 2020). DOI: 10.3389/fimmu.2020.568927. URL: https://hal.inria.fr/hal-03151775.

[32] S. B. Sirima, L. Richert, A. Chêne, A. T. Konate, C. Campion, S. Dechavanne, J.-P. Semblat, N. Benhamouda, M. Bahuaud, P. Loulergue, A. Ouédraogo, I. Nébié, M. Kabore, D. Kargougou, A. Barry, S. M. Ouattara, V. Boilet, F. Allais, G. Roguet, N. Havelange, E. Lopez-Perez, A. Kuppers, E. Konaté, C. Roussillon, M. Kanté, L. Belarbi, A. Diarra, N. Henry, I. Soulama, A. Ouédraogo, H. Esperou, O. Leroy, F. Batteux, E. Tartour, N. K. Viebig, R. Thiébaut, O. Launay and B. Gamain. 'PRIMVAC vaccine adjuvanted with Alhydrogel or GLA-SE to prevent placental malaria: a first-in-human, randomised, double-blind, placebo-controlled study'. In: *The Lancet Infectious Diseases* 20.5 (May 2020), pp. 585–597. DOI: 10.1016/S1473-3099(19)30739-X. URL: https://hal.inria.fr/hal-03161802.

[33] P. Soret, L.-E. Vandenborght, F. Francis, N. Coron, R. Enaud, M. Avalos, T. Schaeverbeke, P. Berger, M. Fayon, R. Thiébaut and L. Delhaes. 'Respiratory mycobiome and suggestion of inter-kingdom network during acute pulmonary exacerbation in cystic fibrosis'. In: *Scientific Reports* 10.1 (27th Feb. 2020), p. 3589. DOI: 10.1038/s41598-020-60015-4. URL: https://www.hal.inserm.fr/inserm-02497484.

[34] E. Turner, L. Yao, F. Li and M. Prague. 'Properties and pitfalls of weighting as an alternative to multilevel multiple imputation in cluster randomized trials with missing binary outcomes under covariate-dependent missingness'. In: *Statistical Methods in Medical Research* 29.5 (May 2020), pp. 1338–1353. DOI: 10.1177/0962280219859915. URL: https://hal.archives-ouvertes.fr/hal-02942173.

[35] G. Vial, N. Gensous, H. Savel, C. Richez, E. Lazaro, M.-E. Truchetet, F. Bonnet, I. Pellegrin, R. Thiébaut, P. Blanco and P. Duffau. 'The impact of clopidogrel on plasma-soluble CD40 ligand levels in systemic lupus erythematosus patients: the CLOPUS phase I/II pilot study: Joint Bone Spine'. In: *Joint Bone Spine* 88.2 (2nd Nov. 2020), p. 105097. DOI: 10.1016/j.jbspin.2020.105097. URL: https://hal.archives-ouvertes.fr/hal-03160699.

[36] H. R. Wagstaffe, G. Susannini, R. Thiébaut, L. Richert, Y. Lévy, Y. Lévy, V. Bockstal, J. N. Stoop, K. Luhn, M. Douoguih, E. M. Riley, C. Lacabaratz and M. R. Goodier. 'Durable natural killer cell responses after heterologous two-dose Ebola vaccination: NPJ Vaccines'. In: *NPJ vaccines* 6.1 (29th Jan. 2021), p. 19. DOI: 10.1038/s41541-021-00280-0. URL: https://hal.archives-ouvertes.fr/hal-03158135.

[37] A. Wiedemann, E. Foucat, H. Hocini, C. Lefebvre, B. Hejblum, M. Durand, M. Krüger, A. K. Keita, A. Ayouba, S. Mély, J.-C. Fernandez, A. Touré, S. Fourati, C. Lévy-Marchal, H. Raoul, E. Delaporte, L. Koivogui, R. Thiébaut, C. Lacabaratz and Y. Lévy. 'Long-lasting severe immune dysfunction in Ebola virus disease survivors'. In: *Nature Communications* 11.1 (24th July 2020), p. 3730. DOI: 10.1038/s41467-020-17489-7. URL: https://www.hal.inserm.fr/hal-03160838.

International peer-reviewed conferences

[38] M. Avalos, H. Touchais and M. Henríquez-Henríquez. 'A decision-making tool to fine-tune abnormal levels in the complete blood count tests'. In: ML4H - Machine Learning for Health workshop at NeurIPS 2020. Vancouver / Virtual, Canada: https://ml4health.github.io/2020/, 7th Dec. 2020. URL: https://hal.inria.fr/hal-03085426.

[39] M. Avalos, H. Touchais and M. Henríquez-Henríquez. 'Optimising criteria for manual smear review following automated blood count analysis: A machine learning approach'. In: WICT 2020 - 10th World Congress on Information and Communication Technologies. Virtual, United Kingdom, 16th Dec. 2020. URL: https://hal.inria.fr/hal-03133078.

Conferences without proceedings

[40] M. Alexandre, M. Prague, Y. Lévy and T. Rodolphe. 'Comparison of AUC in clinical trials with follow up censoring: Application to HIV therapeutic vaccines'. In: ISCB 2020 - 41th Annual Conference of the International Society for Clinical Biostatistics. Krakow / Virtual, Poland, 23rd Aug. 2020. URL: https://hal.archives-ouvertes.fr/hal-03151377.

Scientific books

[41] R. Genuer and J.-M. Poggi. *Random Forests with R*. Use R! 2020, p. 98. DOI: 10.1007/978-3-030-5 6485-8. URL: https://hal.archives-ouvertes.fr/hal-03066152.

Scientific book chapters

[42] R. Azzi, S. Despres and G. Diallo. 'KEFT: Knowledge Extraction and Graph Building from Statistical Data Tables'. In: *Communications in Computer and Information Science*. Vol. 1287. CCIS - Communications in Computer and Information Science. 19th Nov. 2020, pp. 701–713. DOI: 10.1007/978-3-030-63119-2_57. URL: https://hal.archives-ouvertes.fr/hal-03145214.

[43] R. Azzi, S. Despres and G. Diallo. 'NutriSem: A Semantics-Driven Approach to Calculating Nutritional Value of Recipes'. In: *Trends and Innovations in Information Systems and Technologies. WorldCIST 2020. Advances in Intelligent Systems and Computing*. 18th May 2020, pp. 191–201. DOI: 10.1007/978-3-030-45688-7_20. URL: https://hal.archives-ouvertes.fr/hal-031452 23.

[44] B. Xu, C. Gil-Jardiné, F. Thiessard, É. Tellier, M. Avalos and E. Lagarde. 'Pre-Training a Neural Language Model Improves the Sample Efficiency of an Emergency Room Classification Model'. In: *Proceedings of the 33rd International Florida Artificial Intelligence Research Society Conference*. North Miami Beach, United States: https://www.aaai.org/Library/FLAIRS/flairs20conte nts.php, 2020. URL: https://hal.inria.fr/hal-02611917.

Doctoral dissertations and habilitation theses

[45] L. Capitaine. 'Random forests for high-dimensional longitudinal data.' Université de Bordeaux, 17th Dec. 2020. URL: https://tel.archives-ouvertes.fr/tel-03151753.

Reports & preprints

[46] D. Agniel, L. Parast and B. Hejblum. *Doubly-robust evaluation of high-dimensional surrogate markers*. 2020. URL: https://hal.inria.fr/hal-03100499.

[47] L. Capitaine, J. Bigot, R. Thiébaut and R. Genuer. *Fréchet random forests for metric space valued regression with non Euclidean predictors*. 15th Dec. 2020. URL: https://hal.archives-ouverte s.fr/hal-03066146.

[48] C. Colas, B. P. Hejblum, S. Rouillon, R. Thiébaut, P.-Y. Oudeyer, C. Moulin-Frier and M. Prague. *EpidemiOptim: a Toolbox for the Optimization of Control Policies in Epidemiological Models*. 2020. URL: https://hal.inria.fr/hal-03099898.

[49] A. Collin, M. Prague and P. Moireau. *Estimation for dynamical systems using a population-based Kalman filter - Applications to pharmacokinetics models*. 15th June 2020. URL: https://hal.inri a.fr/hal-02869347.

[50] P. Freulon, J. Bigot and B. Hejblum. *CytOpT: Optimal Transport with Domain Adaptation for Interpreting Flow Cytometry data*. 2020. URL: https://hal.inria.fr/hal-03100405.

[51] B. P. Hejblum, K. Kunzmann, E. Lavagnini, A. Hutchinson, D. Robertson, S. Jones and A. Eckes-Shephard. *Realistic and Robust Reproducible Research for Biostatistics*. 2020. DOI: 10.20944/prep rints202006.0002.v1. URL: https://hal.inria.fr/hal-03100421.

[52] V. Philipps, B. P. Hejblum, M. Prague, D. Commenges and C. Proust-Lima. *Robust and Efficient Optimization Using a Marquardt-Levenberg Algorithm with R Package marqLevAlg*. 2020. URL: https://hal.inria.fr/hal-03100489.

[53] M. Prague, L. Wittkop, Q. Clairon, D. Dutartre, R. Thiébaut and B. P. Hejblum. *Population modeling of early COVID-19 epidemic dynamics in French regions and estimation of the lockdown impact on infection rate*. 28th Apr. 2020. URL: https://hal.archives-ouvertes.fr/hal-02555100.

[54] M. Rojas Castro, M. Avalos, B. Contrand, M. Dupuy, C. Sztal-Kutas, L. Orriols and E. Lagarde. *Health conditions and the risk of home injury in French adults: Results from a prospective study of the MAVIE cohort*. 2020. URL: https://hal.inria.fr/hal-03085489.

10.3 Other

Scientific popularization

[55] R. Thiébaut and J. Jongwane. 'Comment la modélisation peut-elle aider au développement des vaccins ?' In: *Interstices* (8th July 2020). URL: https://hal.inria.fr/hal-02929092.

[56] B. Xu, L. Bourdois, C. Gil-Jardine, E. Tellier, F. Thiessard, M. Avalos-Fernandez and E. Lagarde. 'Classification automatique du langage de données du service hospitalier des urgences'. In: 3e Journée Dataquitaine : IA, RO et Data Science. Talence, France, 13th Feb. 2020. URL: https://hal.inria.fr/hal-02615327.

10.4 Cited publications

[57] R. M. Granich, C. F. Gilks, C. Dye, K. M. De Cock and B. G. Williams. 'Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model'. In: *Lancet* 373.9657 (2009). 0140 6736 English, pp. 48–57.

[58] L. Hood and Q. Tian. 'Systems approaches to biology and disease enable translational systems medicine'. In: *Genomics Proteomics Bioinformatics* 10.4 (2012), pp. 181–5.

[59] C. Lewden, D. Salmon, P. Morlat, S. Bevilacqua, E. Jougla, F. Bonnet, L. Heripret, D. Costagliola, T. May and 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* 34.1 (2005). 0300 5771 English, pp. 121–130.

[60] B. Pulendran. 'Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology'. In: *Nature Reviews Immunology* 9.10 (2009), pp. 741–7.

[61] C. Schubert. 'Systems immunology: complexity captured'. In: *Nature* 473.7345 (2011), pp. 113–4.