

RESEARCH CENTRE

**Inria Centre at the University of
Bordeaux**

IN PARTNERSHIP WITH:

Université de Bordeaux, INSERM

2024

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

Inria

Contents

| | |
|--|-----------|
| Project-Team SISTM | 1 |
| 1 Team members, visitors, external collaborators | 2 |
| 2 Overall objectives | 4 |
| 3 Research program | 5 |
| 3.1 Axis 1 - High-Dimensional Statistical Learning | 6 |
| 3.2 Axis 2 - Mechanistic learning | 7 |
| 3.3 Axis 3 - Translational vaccinology and design | 8 |
| 4 Application domains | 9 |
| 5 Social and environmental responsibility | 9 |
| 5.1 Footprint of research activities | 9 |
| 5.2 Impact of research results | 10 |
| 6 Highlights of the year | 11 |
| 6.1 Phase I trial for dendritic cell-based HIV vaccine | 11 |
| 6.2 Impact of vaccination on the COVID-19 Epidemics in France | 11 |
| 6.3 Awards | 11 |
| 6.4 HDR defense | 11 |
| 6.5 Team leader | 12 |
| 7 New software, platforms, open data | 12 |
| 7.1 New software | 12 |
| 7.1.1 CarbPack | 12 |
| 8 New results | 13 |
| 8.1 High-dimensional statistical learning | 13 |
| 8.1.1 Valid inference in high-dimension | 13 |
| 8.1.2 New methods for the analysis of high-dimensional longitudinal data | 13 |
| 8.1.3 Machine learning applied to EHR | 14 |
| 8.1.4 Microbiota data | 14 |
| 8.2 Mechanistic learning | 14 |
| 8.2.1 Response to Ebola vaccination | 14 |
| 8.2.2 Between host modeling of COVID-19 Epidemics | 14 |
| 8.2.3 Analysis of Vaccine trials with analytic treatment interruptions | 14 |
| 8.3 Translational vaccinology and design | 15 |
| 8.3.1 Vaccine trial results and milestones | 15 |
| 8.3.2 Response to COVID-19 vaccine in immunosuppressed populations | 15 |
| 8.3.3 Knowledge transfer | 15 |
| 9 Partnerships and cooperations | 16 |
| 9.1 International initiatives | 16 |
| 9.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program | 16 |
| 9.1.2 STIC/MATH/CLIMAT AmSud projects | 17 |
| 9.1.3 Participation in other International Programs | 17 |
| 9.2 International research visitors | 17 |
| 9.2.1 Visits of international scientists | 17 |
| 9.2.2 Visits to international teams | 18 |
| 9.3 European initiatives | 19 |
| 9.3.1 H2020 projects | 19 |
| 9.3.2 Other european programs/initiatives | 19 |

| | |
|--|-----------|
| 9.4 National initiatives | 21 |
| 9.4.1 Various Partnership | 23 |
| 10 Dissemination | 23 |
| 10.1 Promoting scientific activities | 23 |
| 10.1.1 Scientific events: organisation | 23 |
| 10.1.2 Scientific events: selection | 23 |
| 10.1.3 Journal | 24 |
| 10.1.4 Invited talks | 24 |
| 10.1.5 Leadership within the scientific community | 25 |
| 10.1.6 Scientific expertise | 25 |
| 10.1.7 Research administration | 25 |
| 10.2 Teaching - Supervision - Juries | 25 |
| 10.2.1 Teaching | 25 |
| 10.2.2 Supervision | 26 |
| 10.2.3 Juries | 28 |
| 10.3 Popularization | 29 |
| 10.3.1 Specific official responsibilities in science outreach structures | 29 |
| 10.3.2 Participation in Live events | 29 |
| 10.3.3 Others science outreach relevant activities | 30 |
| 11 Scientific production | 30 |
| 11.1 Major publications | 30 |
| 11.2 Publications of the year | 31 |
| 11.3 Cited publications | 37 |

Project-Team SISTM

Creation of the Project-Team: 2015 January 01

Keywords

Computer sciences and digital sciences

- A3.1.1. – Modeling, representation
- A3.1.10. – Heterogeneous data
- A3.1.11. – Structured data
- A3.3.2. – Data mining
- A3.3.3. – Big data analysis
- A3.4.1. – Supervised learning
- A3.4.2. – Unsupervised learning
- A3.4.3. – Reinforcement learning
- A3.4.4. – Optimization and learning
- A3.4.5. – Bayesian methods
- A5.2. – Data visualization
- 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
- A9.6. – Decision support

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
- B9.5.6. – Data science
- B9.8. – Reproducibility

1 Team members, visitors, external collaborators

Research Scientists

- Melanie Prague [Team leader, INRIA, Researcher, HDR]
- Quentin Clairon [INRIA, Researcher]
- Boris Hejblum [INSERM, Researcher, HDR]

Faculty Members

- Marta Avalos Fernandez [UNIV BORDEAUX, Associate Professor, HDR]
- Robin Genuer [UNIV BORDEAUX, Associate Professor, HDR]
- Edouard Lhomme [UNIV BORDEAUX and UNIV HOSPITAL BORDEAUX, Associate Professor]
- Laura Richert [UNIV BORDEAUX and UNIV HOSPITAL BORDEAUX, Professor, HDR]
- Rodolphe Thiebaut [UNIV BORDEAUX and UNIV HOSPITAL BORDEAUX, Professor, HDR]
- Linda Wittkop [UNIV BORDEAUX and UNIV HOSPITAL BORDEAUX, Professor, HDR]

Post-Doctoral Fellows

- Marie Alexandre [INSERM, Post-Doctoral Fellow]
- Benjamin Hivert [INSERM, from Oct 2024]
- Julien Martinelli [INSERM, from Mar 2024]
- Bastien Reyné [INSERM, Post-Doctoral Fellow, from Sep 2024]
- Corentin Segalas [INSERM]

PhD Students

- Kalidou Ba [UNIV BORDEAUX]
- Antonin Colajanni [UNIV BORDEAUX]
- Sara Fallet [UNIV BORDEAUX, from Oct 2024]
- Thomas Ferte [UNIV BORDEAUX]
- Auriane Gabaut [INRIA]
- Iris Ganser [UNIV BORDEAUX]
- Ange-Marie Gouna [INSERM, from Oct 2024]
- Ariel Oscar Guerra Adames [UNIV BORDEAUX, from Sep 2024]
- Benjamin Hivert [INSERM, until Sep 2024]
- Arthur Hughes [UNIV BORDEAUX]
- Zhe Li [INSERM, from Oct 2024]
- Adrien Mitard De Girardier [INSERM, UNIV PARIS CITE]
- Tran Nam-Ahn [UNIV MCGILL, from Dec 2024]
- Ansh Pal [INRIA]
- Justine Remiat [INSERM]

Technical Staff

- Nicolas Boespflug [INSERM, from Feb 2024]
- Zeinab El Hajj [INRIA, Engineer, from Dec 2024]
- Mélanie Huchon [INSERM]
- Quentin Laval [INSERM]
- Zhe Li [INRIA, Engineer, until Sep 2024]
- Myriam Nafti [INSERM]
- Clément Nerestan [INRIA, Engineer]
- Anton Ottavi [INSERM, Engineer]
- François Plessier [UNIV BORDEAUX, from Sep 2024]
- Anne-Andrée Ruiz [INSERM, from Sep 2024]
- Hélène Savel [UNIV BORDEAUX, Engineer]
- Alice Simon [INSERM, from Nov 2024]
- Panthea Tzourio [INSERM, Engineer]

Interns and Apprentices

- Hugo Alves [UNIV BORDEAUX, Intern, from Jun 2024 until Sep 2024]
- Michael Burtin [INRIA, Intern, from Feb 2024 until Jul 2024]
- Chloé Dumas de la Roque [UNIV BORDEAUX, Intern, from Oct 2024]
- Arthur Le Roudier [UNIV BORDEAUX, Intern, from Jun 2024]
- Chloé Renault [INSERM, Intern, from Jun 2024 until Aug 2024]
- Xi Jin Susie Wang [INSERM, Intern, from Apr 2024 until Sep 2024]

Administrative Assistants

- Ellie Correa Da Costa De Castro Pinto [INRIA]
- Sandrine Darmigny [INSERM]

Visiting Scientists

- John Barrera [Universidad de Valparaíso, Chile, from Sep 2024 until Oct 2024, PhD Student]
- Susana Eyheramendy [Universidad Adolfo Ibáñez, Chile, from Mar 2024 until Mar 2024, Professor]
- Cristian Meza [Universidad de Valparaíso, Chile, from Mar 2024 until Mar 2024, Professor]
- Lander Rodríguez [Basque Center for Applied Mathematics (BCAM), Spain, from Apr 2024 until Jul 2024, PhD Student]

External Collaborator

- Lucie Bourguignon [INSERM, from Sep 2024]

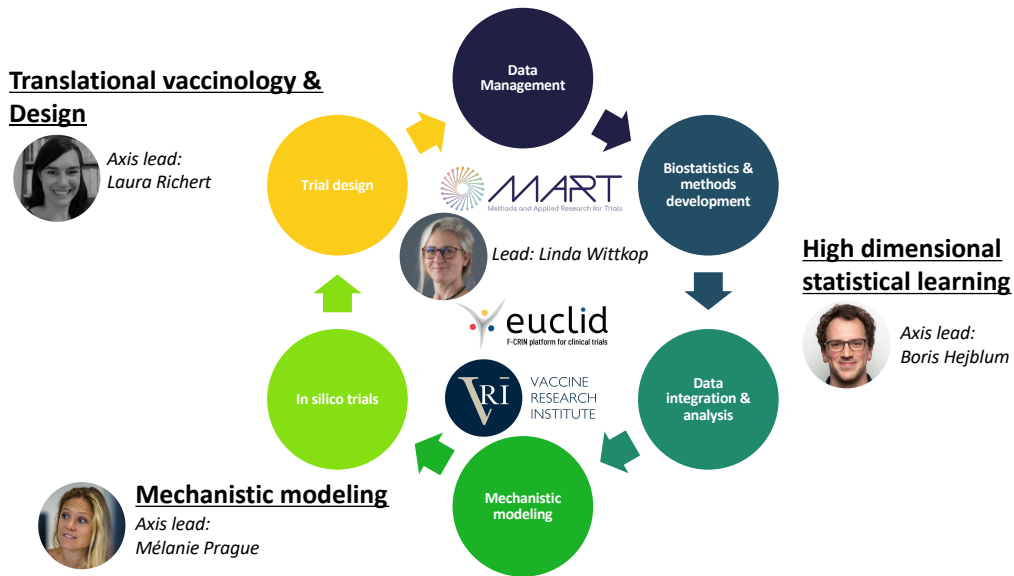


Figure 1: The SISTM wheel. Presentation of the three axes.

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 high dimensional data in small sample size studies.

The methods developed are relevant in many other applications beyond those encountered in the SISTM team. However, the focus devoted to vaccine development is justified by its importance from a public health perspective, and a long-standing expertise in this application field that maximizes the relevance and implementation of the methods developed. This equilibrium between the methodological and applied work reached over the last years is a fundamental motivation for each member of the SISTM team, regardless of complementary backgrounds across researchers (from applied mathematics to public health). This equilibrium is maintained by the organization of the team as well as the collaborations established especially through the Vaccine Research Institute, Bordeaux University, Inserm and Inria. Thus, we are able to collaborate for the development of new methods, and also to translate our innovations (either new analytical methods or applied results) to clinicians and immunologists – first in our collaborative networks, and then beyond. Figure 1 illustrates this synergy and materializes the three research axis of the team: high dimension statistical learning, mechanistic modelling, and translational vaccinology and design.

Biological and clinical research has dramatically changed thanks to technological advances, leading to the possibility of measuring many more biological parameters than previously thanks to high-throughput methods. Clinical research studies can now include traditional measurements such as clinical status, but also (tens of) thousands of cell populations, peptides, gene expressions, etc. for a given participant at a single time point. This has facilitated knowledge transfer from basic to clinical science (from “bench to bedside”) and vice versa, a process often called “translational medicine”. However, the analysis of these large amounts of data requires specific methods, especially to obtain a global understanding of the information inherent to complex systems through an “integrative analysis”. Systems like the immune system are complex because of the many interactions within and between several scales (within cells, between cells, in different tissues, between individuals, between various species). This has led to a new field called “Systems biology” rapidly adapted to specific topics such as “Systems Immunology”

[98], “Systems vaccinology” [95], “Systems medicine” [79]. From the statistical point of view, two main challenges arise: i) to adequately deal with the massive amount of data, and ii) to find relevant models capturing observed data.

First, with respect to the relatively moderate number of participants in vaccine studies and clinical trials, this profusion of high-throughput “omics” data often sets us in a ultra high-dimension context. This mandates updated statistical tools able to tackle this wealth of information. On top of the challenge signal extraction and dimension reduction, there is a redundancy of the information across data modalities, that in turn can be leveraged to boost statistical methods and harness artificial intelligence approaches to predict immunological surrogate endpoints from early indicators.

Second, once a small amount of markers has been selected, we use modeling approaches to understand the biological mechanism (specifically in vaccinology antibodies kinetics or viral dynamics [91]). In our work we are interested in the inverse problem: how can we infer the mechanism of a biological process from data. It can be modeled using differential equations (mainly ordinary but could extend to partial and stochastic). The challenge in our methods rely in the type of collected data which are sparse (as opposed to measured in continuous time), with measurement error and repeated across multiple individuals. Thus, we adopt nonlinear mixed-effects model population approach [84]. Construction of these models is a challenging process which requires confirmed expertise, advanced statistical methods and the development of software tools.

Finally, once a model has been defined and validated, it is possible to perform *in silico* trials to predict further strategies. In particular, a systems personalized vaccinology approach [92] using multidimensional immunogenicity data from clinical trials and statistical models (such as optimal control or reinforcement learning) can help improve the selection of optimized vaccine strategies that can then be tested again in subsequent clinical trials.

Domains of application of our methods in vaccinology focuses on, but not limited to, Ebola virus, Human Immunodeficiency Virus (HIV) virus and SARS-CoV-2 virus. The choice of these applications is deliberate and important for the relevance of the results and their translation into practice, thanks to a longstanding collaboration with several immunology research teams and the implication of the team in VRI - the Labex Vaccine Research Institute.

The SISTM team benefits from a very rich ecosystem (also represented in part in the figure 1). Firstly, it is one of the rare teams belonging to both 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 either Inserm teams or Inria teams from the two distinct 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 (UMS 54 MART directed by LW), from Bordeaux Hospital (USMR directed by LR and previously by RT), from F-CRIN (Euclid platform, directed by LR and EL), from the international consortia linked to the Vaccine Research Institute (for which SISTM is leading the data science division). Finally, the team is very much involved in teaching activities at Bordeaux University and ISPED Institute, especially through the Graduate’s program Digital Public Health (directed by RT) and the Master of Public Health (first year in e-learning led by MA, Biostatistics led by RG and Public Health Data Science led by RT). A better description of all these interaction can be found in section Teaching (10.2) and section Fundings (9).

In term of positioning in regards of other teams at Inria and in France, the application domain (immunology and vaccine development) is nearly unique with the exception of DRACULA in Lyon. DRACULA like other teams at Inria (MONC, CARMEN, M3DISIM) or Inserm (IAME) or international groups (e.g. A. Perelson lab in Los Alamos, Schiffer lab in Fred Hutchinson Cancer center) are also developing mathematical models but rarely with the integration of high dimensional data. In other hand, groups such as Raphael Gottardo lab in Lausanne (previously at the Fred Hutchinson in Seattle) are developing methods for high dimensional data in immunology but are not using dynamical models.

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 and design (leader Laura Richert).

3.1 Axis 1 - High-Dimensional Statistical Learning

The specific objectives are:

- To unlock the analysis of high-dimensional longitudinal data by developing suitable statistical approaches, in particular for applications to longitudinal high-throughput data (e.g. microbiome, transcriptome, cytomics) generated in vaccine trials.
- To leverage prior biological knowledge and formally incorporate it into statistical models to tackle the small n large p setting, one of the characteristics of early phase vaccine trials.
- To advance adaptive clustering methods of high-dimensional data in both supervised and unsupervised settings, especially to infer the proportions of cellular population from gene expression measurements and also to identify gene whose expression is key in segmenting transcriptomic measurements across vaccine arms or disease severity for instance.
- To perform feature selection and dimension reduction of high-dimensional molecular and cellular data, as a first step to feed such information into mechanistic models.

Despite being high-dimensional, biomedical data from high-throughput technologies is rarely analyzed in its entirety due to its size or its complexity. For example, in cellular phenotyping data, only a limited number of markers are used to quantify a pre-defined set of cell types; this strategy precludes the discovery of new cell types defined by new combinations of markers. This issue is exacerbated by mass cytometry technologies, which enable the measurement of up to 100 markers on a single cell.

However, measuring specific cells across a large number of intracellular and surface markers requires substantial amounts of blood, ideally fresh, making it difficult to implement such measurements on large sample sizes with multiple repeated measurements. This motivates the exploration of replacing cell phenotyping with transcriptomics analysis in whole blood, as gene expression can be measured more easily and frequently with a much finer temporal resolution (using finger prick at-home self-sampling technology [101]). This ambitious endeavor goes beyond previous work done on this topic using standard deconvolution approaches [81]. By using more sophisticated statistical [72], machine learning [22], and artificial intelligence [104] models (in particular for adaptive clustering, robust to unobserved cell populations), by exploiting public databases of cytometry data coupled newly available single cell transcriptomics measurements, and by explicitly leveraging the repeated aspect of longitudinal observations from vaccine trial measurements, we set ourselves to successfully study and develop methods delivering accurate cell proportions estimates from gene expression data.

In addition, among high-throughput omics data, the microbiome is also becoming an increasingly important component in understanding the immune system[88]. The compositional nature of these data, along with their hierarchical phylogenetic structure particularly suited to tree-based models, coupled with their high-dimension requires the use of adequate statistical tools [100].

Furthermore, while those high-throughput molecular and cellular data have an unquestionable value for diving into underlying mechanisms governing and deepening our understanding of the human immune system, we want to determine whether they could be used as early surrogate markers for correlates of protection in vaccine studies (such as antibody titers after vaccination). Due to their high-dimensional nature, answering this question requires the development of new mediation approaches [69] to develop this emerging field of vaccinomics epidemiology.

Outside biological data generated in clinical trials, electronic health records from hospital data warehouse systems are also representing an opportunity for studying infectious diseases and requires specific approaches. Several works have been done on this topic in the SISTM team [76, 78, 99, 107, 77].

Regarding this research axis, there are some common interest with other Inria teams such as **HeKA**, **Soda**, and **PreMeDICAL** in regards of the use of machine learning approaches applied to medical data or **Mind** and **Aramis** that are more focus on brain applications. Applications in SISTM are focused on analyzing high-throughput omics data (nearly no imaging) in immunology and vaccine trials. Also, modeling biological networks as done in **Beagle** or **Dyliss** Inria teams is not an objective of SISTM, the

data recorded in human clinical trials being unsuited because of their sparsity. At the international level, the main competitors are groups engaged in biostatistical methods development for the analysis of omics data such as [Jeff Leek](#) (previously at John Hopkins, now at Fred Hutchinson, Seattle), [Raphael Gottardo](#) (previously at Fred Hutchinson, Seattle, now at Université de Lausanne, Switzerland) or [Mark Robinson](#) (Prof at the University of Zurich, Switzerland).

3.2 Axis 2 - Mechanistic learning

The specific objectives are:

- To develop methods for statistical inference of differential equations model parameters in population framework.
- Within-host modeling of immunological and virological dynamics in samples of individuals.
- Between-host modeling of dynamics of epidemics in populations.
- Use mechanistic model as in silico platform for exploration of counterfactual scenarios with application in implementing control strategies toward personalized medicine.

When studying the dynamics of some given markers one can for instance use descriptive models summarizing the dynamics over time in term of slopes of the trajectories [103]. These slopes can be compared between treatment groups or according to patients' characteristics. Mechanistic modeling, that is dynamical models based on Ordinary Differential Equations (ODE), could be preferred as it integrates knowledge about the biological mechanism and it carries causal interpretation of the observed phenomenon [68, 94]. Thus, in this axis, we focus on inference of model parameters of mechanistic models in population of subjects (e.g. from a clinical trial). This modeling is constituted by three features: 1/ a dynamical model, which describes a phenomenon, often based on ODE (but also possibly partial and stochastic DE) 2/ a statistical model, which describes the variability that exists in data and the heterogeneity between individuals, and 3/ an observational model, which relates what is observable with error in the mathematical model.

The definition of the model needs to identify the parameter values that fit the data. Contrary to Inria team such as MAKUTU or BEAGLE, which are interested in simulation scheme for large differential equation systems, we focus on inverse problems for inference of parameters from data. In clinical research, this is challenging because data are sparse, and often unbalanced, coming from populations of individuals. A substantial inter-individual variability is always present and needs to be accounted as this is the main source of information. Many approaches have been developed to estimate the parameters of non-linear mixed models (NLME) including Bayesian approach [106], semi-parametric approaches [105] or penalized likelihood approach (in house NIMROD program [93]). The SAEM algorithm [83], as implemented in Monolix [97], is now also used for many of our projects. We however, continue to participate in the development of related methods in collaboration with ex - Inria team XPOP. We also devote a large part of this axis methodological research to the development of alternative methods for estimation in NLME-ODEs models.

From a computational perspective, the stochastic approximation of the EM algorithm (SAEM) provides accurate estimations for medium-sized parametric NLME-ODEs. For high-dimensional settings, alternative approaches to SAEM, such as those based on variational inference [82], have been proposed for generalized linear mixed models [89]. However, these methods have not yet been extended to NLME-ODEs. In the context of semi-parametric inference of ODEs, the universal approximation property of neural networks (NNs) has justified their use as proxies for missing model structures [108]. Nevertheless, this is usually limited to single-subject settings. While some studies have begun to consider population contexts [96, 86], these approaches remain inadequate for sparse data scenarios. A great amount of this axis work now focuses on estimation methods using concepts/devices coming from NNs, variational inference and inverse problem regularization, to construct high-dimensional, semi-parametric and properly regularized inference methods for mechanistic models, in the vein of hybrid modeling.

The integration of ordinary differential equation (ODE) models in our work enables a detailed examination of within-host and between-host dynamics of infectious diseases. At the within-host level, ODE models describe the interactions between pathogens and host immune responses, such as viral replication and immune clearance. These models provide insights into mechanisms like virus propagation and immune cell dynamics, as demonstrated for example in studies on HIV [102], Ebola [90, 71] and SARS-CoV-2 [70, 73]. Regarding the between-host dynamics, we extensively described the COVID-19 pandemics inferring the effect of vaccination and non-pharmaceutical interventions [75, 26].

Having a good mechanistic model with a population approach in a biomedical context 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 biological mechanisms rather than simple regressions. Control theory (Inria team ASTRAL), game theory (Inria team SCOOL) and learning approaches (Inria team FLOWERS) may serve for defining optimal interventions or optimal designs to evaluate new interventions. We made a proof of concept of such open-loop control problem in the within-host setting. We model the response to Interleukin-7 (IL-7) injections in HIV-infected patients, and that has allowed to design new trials finally implementing personalized medicine [80]. We also made a proof of concept of such open-loop control problem in the between-host setting. [74]. We introduced *EpidemiOptim*, a Python toolbox designed to optimize epidemic control policies through the integration of epidemiological models and machine learning algorithms, including reinforcement learning and evolutionary algorithms. The toolbox's utility is demonstrated through a case study optimizing COVID-19 lockdown policies, balancing health outcomes and economic impacts using a Susceptible-Exposed-Infectious-Removed (SEIR) model fitted to French data. We still devote a large part of this axis methodological research to the development of methods around personalized medicine and targeted numerical public health.

Regarding this axis, the SISTM team compares to DRACULA, BIOCORE, MONC and COMPO Inria team. However, differences arise in two ways 1/ the application field is immunology, vaccinology and infectious diseases and 2/ we adopt a population approach. This last point results in using simpler models in which it is possible to infer parameters from sparse data by taking advantage of an underlying mechanism common to all patients. Regarding the modeling, our international competitors and collaborators are Perelson's lab in Los Alamos USA and Schiffer's lab in Fred Hutchinson USA. Finally, our work on *in silico* simulation is closely related to a digital twin of clinical trials. In this sense, it can be compared to the work conducted by the SIMBIOTX team at Inria.

3.3 Axis 3 - Translational vaccinology and design

The specific objectives are:

- To accelerate the vaccine development by in depth analysis of data generated in early clinical trials and
- designing the next trials with development of new adaptative designs and *in silico* trials in collaboration with immunologists and clinicians.

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, translational trials), the number of possible candidate vaccine strategies against a given pathogen that needs to be down-selected is potentially very large. Moreover, during early clinical development there are most often no validated surrogate endpoints to predict the clinical efficacy of a vaccine strategy based on immunogenicity results that could be used as a consensus immunogenicity endpoint and down-selection criterion. This implies considerable uncertainty about the interpretation of immunogenicity results and about the potential value of a vaccine strategy as it transits through early clinical development. Given the complexity of the immune system and the many unknowns in the generation of a protective immune response, early vaccine clinical development nowadays thus takes advantage of high throughput (or "omics") methods allowing to simultaneously assess a large number of response markers at different levels ("multi-omics")

of the immune system. Outside of the context of emergency vaccine development during a pandemic, this has induced a paradigm shift towards early-stage and translational vaccine clinical trials including fewer participants but with thousands of data points collected on every single individual. This is expected to contribute to acceleration of vaccine development thanks to a broader search for immunogenicity signals and a better understanding of the mechanisms induced by each vaccine strategy. However, this remains a difficult research field, both from the immunological as well as from the statistical perspective. Extracting meaningful information from these multi-omics data and transferring it towards an acceleration of vaccine development requires adequate statistical methods (in close collaboration with axis 1), state-of-the-art immunological technologies and expertise, and thoughtful interpretation of the results.

Our main current areas of application here are early phase trials of HIV and Ebola vaccine strategies, in which we participate from the initial trial design to the final data analyses. We are also involved in the development of next-generation pan-Coronavirus vaccines.

Research on novel trial designs for early phase vaccine trials is carried out by the team within PEPR Santé Numérique SMATCH, and with PhD thesis (such as the Inserm-Inria funded thesis on multi-armed bandit algorithms for vaccine trials by Cyrille Kone; co-supervisor E Kaufmann Inria Lille).

In regards of the number of trials we are dealing with, the complexity of the data (including clinical and biological high dimensional data), the need for a collaborative tool for data sharing that is respectful of GDPR and health data protection, we have set up a data warehouse system based on the Labkey solution (also used for the Immunespace funded by the NIH). We are currently plugging in our data analysis and data visualization tools. This solution may constitute a very nice way to boost our collaborations but also to facilitate the access to the statistical tools we have developed.

To our knowledge, our specific application to vaccine trials is unique in France. Although some research teams have sometimes applications in this field (e.g. clinical epidemiology team at Inserm U1018 or Inria DRACULA team), there are less devoted to it. Internationally, the closest group to SISTM research axis 3 is the vaccine and infectious disease division of the Fred Hutchinson Institute (Seattle). There are also several groups working on systems immunology mainly in United States such as Mark Davis at Stanford University, Bali Pulendran at Emory University, Rafick Sekaly at Case Western Reserve University, Galit Alter at the Ragon Institute. There are all immunologists integrating bioinformaticians in their groups therefore they are more applying than developing new methods. We have collaborated with several of these groups.

4 Application domains

The main application domain is the clinical immunology of infectious diseases and more specifically vaccine development.

The main infectious diseases concerned up to now are:

- Human Immunodeficiency Virus (HIV);
- Ebola virus (following the 2014 epidemics);
- SARS-Cov2 virus;
- Hepatitis B virus;
- NIPAH virus;

This is not a closed list and new studies are currently settled on other infectious agents (e.g. tuberculosis, Human Papilloma Virus...).

5 Social and environmental responsibility

5.1 Footprint of research activities

National and international programs

- **Coordination of the response to the Referral for primary care clinical research in France - Ministry of Health (September 2021 - April 2022):** The objective was to make proposals to anticipate the implementation of future ambulatory trials in response to an emerging infectious disease and enable them to reach their recruitment targets quickly, and to structure research in primary care more broadly. The response includes a national and international review of COVID-19 ambulatory research and 20 proposals on research strategy, its structuring and the removal of budgetary and regulatory constraints.
- **Participation in Delphi consensus groups:** The objective was to extend the CONSORT and SPIRIT recommendations. Participated in the elaboration of SPIRIT/CONSORT Extension for Surrogate endpoints (2023)
- Laura Richert is the coordinator of the working group "Greener Clinical Research" (Décarbonation de la recherche clinique) within the Recap/Inserm network. She is also a member of the "Greener Trials" network (MRC, UK) and a member of the "Sustainable Development" working group of the CNCR. Thomas Ferte has developed an R package called CarbPack R, designed to facilitate the estimation of the carbon footprint of statistical analyses in R on a local computer. The package serves as a wrapper for the [Green Algorithm calculator](#).

5.2 Impact of research results

Drug licensure and patents

- Participant as "Inventor" (Décret n°96-858 du 2 octobre 1996) to the development and the authorization for commercialization (1/7/2020) of the Janssen Zabdeno® (Ad26.ZEBOV) and Mvabea® (MVA-BN-Filo) vaccines against Ebola virus infection.
- Patent 20 306 527.1 on "Use of CD177 as biomarker of worsening in patients suffering from COVID-19" (10/12/2020)
- Participant as "Inventor" (1/7th) for patent WO2021058914A1/FR1910515 on "Prediction of the content of omega-3 polyunsaturated fatty acids in the retina by measuring 7 cholesterol ester molecules"

Public/Private partnership

- In the context of clinical trials: Johnson and Johnson (IMI-2 Anti-Ebola vaccine trial Ebovac and Prevac; Merck (Anti-Ebola vaccine trial Prevac/Prevac-up); Iliad Biotechnologies (Anti-pertussis vaccine trial BPZE-1); Gilead Sciences (IP-Cure-B)
- In the context of CIFRE PhD funding: Ipsen (LR HS, 2020-2023). Thesis defended in 2023.

Multicenter clinical trials on vaccine research

- Coordination clinical trials through the Euclid/F-CRIN, CIC1401 platform: Leading Phase II international clinical trials (steering and methodology) for projects BPZE-1, Ebovac2, IP-Cure-B, Prevac, Prevac-Up et PrimalVac (see fundings section).
- Methodology for clinical trials:
 - International phase II anti-Ebola vaccine trial PREVAC (NCT02876328) and EDCTP2 PREVAC-UP
 - International phase I anti-Malaria vaccine trial PRIMALVAC (NCT02658253)
 - French Phase I/II anti-HIV vaccine trial ANRS VRI01 (NCT02038842)
 - French Phase I anti-HIV vaccine trial ANRS VRI06 (NCT04842682)
 - Monocenter anti-pertussis phase I vaccine trial BPZE-1 (NCT02453048)

- French phase II anti-pneumococcal vaccine trial PNEUMOVAS (NCT03069703)
- French phase II anti-pneumococcal vaccine trial SPLENEVAC2 (NCT03873727)
- French phase II anti-meningococcal vaccine trial SPLENMENGO (NCT04166656)
- French phase II anti-HPV vaccine trial PRIMAVERA (NCT01687192)
- French Phase IV anti-Dengue vaccine trial (LR, trial set-up ongoing)
- Cohort study of anti-COVID-19 vaccination in specific populations (ANRS0001S COV-POPART)
- Cohort study of HIV infected patients in Nouvelle-Aquitaine (ANRS CO3 Aquitaine)
- Cohort study of HIV-2 infected patients in France (ANRS CO5 VIH-2)
- Cohort study of co-infected patients with HIV and Hepatitis in France (ANRS CO13 HEPAVIH)
- International phase II proof of concept trial IP-cure-B . Educating the liver immune environment through TLR8 stimulation followed by NUC discontinuation. (ANRS HB 07 IP-Cure-B Trial)
- French phase I anti-SARS-COV2 nasal vaccine trial MUCO-BOOST.

6 Highlights of the year

6.1 Phase I trial for dendritic cell-based HIV vaccine

The VRI06 phase 1 trial, conducted in France and Switzerland from May 2021 to October 2022, assessed the safety and immunogenicity of the CD40.HIVRI.Env vaccine, targeting the gp140 Env from HIV Clade C 96ZM651, in 72 HIV-negative volunteers. The trial, randomized at a 5:1 ratio between active and placebo groups, tested different doses of the CD40.HIVRI.Env vaccine, adjuvanted with Hiltonol®, alone or in combination with the DNA-HIV-PT123 vaccine. Participants were monitored for safety and immune responses until week 48. The study concludes that the CD40.HIVRI.Env vaccine, with or without DNA-HIV-PT123, is safe and effectively induces sustained cellular and humoral immune responses, suggesting its potential usefulness in prime-boost vaccine strategies targeting HIV. This trial has been led by the Vaccine Research Institute for which SISTM is the data division core. This has been published in *Lancet EClinicalMedicine* [33].

Research on novel trial designs for early phase vaccine trials is carried out by the team within PEPR Santé Numérique SMATCH, and with PhD thesis (such as the Inserm-Inria funded thesis on multi-armed bandit algorithms for vaccine trials by Cyrille Kone; co-supervisor E Kaufmann Inria Lille).

6.2 Impact of vaccination on the COVID-19 Epidemics in France

We constructed a dynamic model of SARS-CoV-2 to assess interventions against COVID-19. Lockdowns and curfews significantly curbed the spread of COVID-19 in France. The deployment of vaccines prevented 160,000 deaths in France by the end of the study period. Had vaccines been available 100 days earlier, an additional 70,000 lives could have been saved in France. This has been published in *Epidemics* [26].

6.3 Awards

The paper [27] was awarded Best Proceedings Paper in the “Impact and Society: Policy, Public Health, Social Outcomes, and Economics” track at the Machine Learning for Health (ML4H) Symposium, organized by the Association for Health Learning and Inference (AHLI).

6.4 HDR defense

Mélanie PRAGUE defended her HDR on "Mechanistic and Statistical Models for Treatment and Control of Infectious Diseases" [60]. This work details the development of mechanistic models, their application to intra-host viral dynamics, and population-level interventions against epidemics, targeting diseases such as HIV, Ebola, and SARS-CoV-2.

Boris Hejblum defended his HDR on "Statistical methods for leveraging high-dimensional data from high-throughput measurements in vaccine clinical development" [58]. The focus is on the methodological challenges and advancements in handling complex data sets typical in vaccine trials, which include not only gene expression studies but also flow-cytometry data for cellular phenotyping, ultimately aiming to enhance our understanding of immune responses and improve vaccine design and efficacy.

6.5 Team leader

Rodolphe Thiébaud took the head of the Bordeaux Population Health research centre that is composed by 10 Inserm research teams including SISTM. Mélanie Prague took the head of the SISTM team.

7 New software, platforms, open data

7.1 New software

7.1.1 CarbPack

Keywords: Climate change, Carbon footprint, Statistical analysis, Greenhouse gas emissions

Scientific Description: Background and objective(s): Greenhouse gas emissions are a major concern worldwide, given their impact on climate change, which in turn threatens human health. Reducing CO₂ emissions (carbon footprint) is key to alleviate greenhouse gas emissions. Scientific research activities themselves generate CO₂ emissions, and a key professional activity of biostatisticians is the conduct of statistical analyses. Although tools to estimate the carbon footprint of statistical code have been developed in Python and as online calculators, an easy-to-use routine tool for clinical epidemiology units using R is lacking. We aimed at implementing the method available in the online calculator "Green algorithms" (www.green-algorithms.org; Lannelongue et al. Adv. Sci. 2021) into the R software for embedded use for analyses computed with R. Material and Methods: We developed the CarbPack R package to estimate the carbon footprint of statistical analyses in R. The package serves as a wrapper for the Green Algorithm calculator and operates through two primary functions. The first function, called at the beginning of the analysis script, identifies the hardware used for computation and starts a timer. The second function, called at the conclusion of the script, stops the timer and calculates both the energy consumption and the carbon footprint. This estimation is based on computation duration, energy consumption of key hardware components, and the carbon intensity (French emission factors 2024) of the electricity used. The tool runs concurrently with the computational program and provides an automatic output of the carbon footprint of the analysis. We tested the package on the statistical analyses of four previously analyzed clinical studies according to their pre-defined statistical analysis plan: a prognostic cohort study (NCT02821208), a diagnostic delayed-type cross-sectional study (NCT03695861), a monocenter randomized clinical trial (NCT04113187), and a before-after cohort study (NCT02411669). All studies involved classical statistical analysis methods including descriptive analysis, regression model and linear mixed models. The package is available on Windows and Linux operating systems. Results: The R package, available on GitHub (<https://github.com/thomasferte/CarbPack>), was used to estimate the carbon footprint of the statistical analyses. The evaluations were conducted on a system with a 6-core AMD Ryzen 5 7530U processor with Radeon Graphics in France. Execution times for a single run of the full analysis per study ranged from 22.7 seconds to 373.29 seconds. Estimated energy consumption varied between 0.13 Wh and 2.17 Wh, with corresponding carbon footprints from 0.01 g CO₂e to 0.15 g CO₂e. For contextualisation, driving 1 kilometer by car emits approximately 100 g CO₂e, the carbon footprint of a portable computer is approximately 350 kg CO₂. Conclusion: We developed an R package that provides a straightforward method for estimating the carbon footprint of statistical code on the local computer. This tool is user-friendly, requiring only three lines of code, and offers a rough estimate that can help raise awareness and encourage users to optimize their computational workflows. While the evaluated studies involved relatively simple statistical analyses and low carbon footprints, more computationally intensive tasks, such as those employing extensive resampling methods like bootstrapping, would likely result in higher footprints. Additionally, the reported

values represent a single execution, whereas, in practice, multiple iterations are often required to produce the final statistical report. More comprehensive evaluations of the environmental impact of academic clinical epidemiology units are warranted, including the impact of the acquisition of the computer equipments and of data storage.

Functional Description: This package provides an R wrapper for the Green Algorithms, enabling users to quantify the carbon footprint of their computations directly within R.

URL: <https://github.com/thomasferte/CarbPack>

Contact: Thomas Ferté

Participants: Thomas Ferte, Laura Richert, Adam Loffler

8 New results

8.1 High-dimensional statistical learning

8.1.1 Valid inference in high-dimension

Participants: Boris Hejblum, Rodolphe Thiébaud.

Statistical inference in high dimension remains challenging. In particular, the sound analysis of transcriptomic data requires valid statistical testing procedures with well calibrated control of the type I error. In addition, the analysis of high-dimensional longitudinal and time-to-event data bears its own layer of complexity. Another example of the complexity of the analysis of such data is the so-called “double dipping”, i.e. using the same data twice, first to identify clusters and second to identify statistically significant variables differentiating those clusters. This leads to poor statistical performances overall without carefully adjusting for this double step – yet it is widely used in transcriptomic data analysis for instance. We developed three new strategies for valid post-clustering testing in low dimension [30]. In addition, two approaches were recently introduced as candidate that could solve this post-inference clustering issue more generally in certain context, namely data-fission [85] and data-thinning [87]; and we characterized their limits showing that they cannot be applied in practice for analyzing gene expression data with data-driven sub-groups in immunological studies [61].

8.1.2 New methods for the analysis of high-dimensional longitudinal data

Participants: Robin Genuer, Boris Hejblum, Rodolphe Thiébaud.

In many health studies, we have to deal with high-dimensional longitudinal data, meaning that in addition to have a very large number of variables at hand, those are repeatedly measured over time. As standard machine learning methods are most of the time tailored for independent observations, new approaches have to be developed to adapt to this new framework and use as much as possible all the available information. We introduced several methods to deal with this kind of data: either we see the different measurements of a variable as an individual curves and generalize random forests by using distances between curves leading to Fréchet random forests [22]; secondly in [43] we consider those repeated measurements as function and study the behavior of functional principal component analysis in presence of missing data in [43]; and finally we leverage reservoir computing technique to forecast short high-dimensional time series in [56].

8.1.3 Machine learning applied to EHR

Participants: Marta Avalos, Boris Hejblum, Rodolphe Thiébaud.

Given the recent advances in machine learning, notably thanks to deep-learning and large language models, we are questioning the impact and interest of such approaches for biomedical research and training [44], but also considering their use and added value for leveraging EHR data [27, 56].

8.1.4 Microbiota data

Participants: Marta Avalos.

Recently, the team has been interested in microbiota data analysis. One communication in international conferences showcased the challenges met [51].

8.2 Mechanistic learning

8.2.1 Response to Ebola vaccination

Participants: Mélanie Prague, Edouard Lhomme, Rodolphe Thiébaud, Linda Wittkop, Laura Richert.

The major focus of the year was the analysis of the data generated within the PREVAC and EBOVAC series project. First, we better characterize the proliferation of CD8+ T cells and inflammation after Ad26.ZEBOV, MVA-BN-Filo vaccination [32]. Second, the five-year follow-up of the PREVAC trial demonstrates that Ad26-MVA and rVSV Ebola vaccine regimens induce sustained long-term T-cell responses, with Ad26-MVA showing superior polyfunctional T-cell activity, and highlights correlations between T-cell and IgG responses [49]. Third, we found no significant association between baseline immune markers of helminth exposure and the antibody response to the Ad26.ZEBOV, MVA-BN-Filo Ebola vaccine regimen, despite lower levels of certain inflammatory and activation markers in participants with helminth exposure [19]. Finally, we showed that antibody concentrations were higher in children than adults one year post-vaccination, with variations also influenced by geography, pre-vaccination antibody levels, and sex [46]. Our findings can guide booster vaccination recommendations and help identify populations likely to benefit from revaccination.

8.2.2 Between host modeling of COVID-19 Epidemics

Participants: Mélanie Prague, Rodolphe Thiébaud.

We investigated the effectiveness of non-pharmaceutical interventions (NPIs) implemented in France during the COVID-19 pandemic. We clearly demonstrated the positive effect of vaccination on the epidemics dynamics [26]. This is one of the highlight of the year.

8.2.3 Analysis of Vaccine trials with analytic treatment interruptions

Participants: Mélanie Prague, Rodolphe Thiébaud, Linda Wittkop, Laura Richert.

Analytical treatment interruption (ATI) is used in HIV research to evaluate new therapeutic strategies but often involves short phases and strict ART restart criteria, potentially limiting the observation of viral setpoints. This study analyzed viral dynamics from 235 individuals across three trials, focusing on time- and magnitude-related virological criteria. Results showed strong correlations between the time-averaged area under the curve (nAUC) and setpoints, with nAUC emerging as an optimal surrogate endpoint. Recommendations highlight the value of ATI phases longer than 12 weeks with regular monitoring and a restart criterion of 10,000 copies/mL to balance patient safety and data quality [16]. This result has important implication for the design of trial within our EHVA consortium.

8.3 Translational vaccinology and design

8.3.1 Vaccine trial results and milestones

Participants: Rodolphe Thiébaud, Laura Richert.

The ANRS VRI06 first-in-human phase I trial of a novel vaccine concept targeting dendritic cells (here as HIV vaccine) has completed all W48 follow-up visits and results are available [33]. We have prepared the designs of subsequent clinical trials of this vaccine candidate as well as of other candidates based on the same platform (HPV vaccine, SARS-COV2 and Pansarbeco vaccines ; Nipah vaccine). This is one of the highlight of the year.

The protocol has been amended with an additional late boost vaccine administration (after W48 since the first injection). Results of the late boost administration will be presented at the CROI conference 2025.

8.3.2 Response to COVID-19 vaccine in immunosuppressed populations

Participants: Mélanie Prague, Linda Wittkop.

Linda Wittkop is involved as a principal investigator in the COV-POPART cohort. The cohort was established to study the immune response to Covid-19 vaccination and its persistence in individuals with immune disorders, with a particular focus on characterizing vaccine failures (immunological and virological). A total of 6,112 adults affected by 10 different pathologies are participating in the study. The statistical analysis of the data generated has started.

Loubet et al [34] compared humoral responses to mRNA COVID-19 vaccination between people with HIV (PWH) on ART and HIV-negative individuals. PWH had lower seroneutralization titers and anti-Spike IgG levels than controls at 1 and 6 months post-vaccination, particularly those with CD4 counts <200/mm³. However, after a booster dose, humoral responses at 12 months were comparable between PWH and controls, even in those with initially low CD4 counts.

Chalouni et al [23] found that higher levels of anti-Spike IgG antibodies were associated with a lower risk of SARS-CoV-2 infection in the control group but not in specific populations with chronic diseases. In specific populations, the predictive performance of anti-Spike IgG, anti-RBD IgG, and neutralizing antibodies for SARS-CoV-2 infection was moderate. Overall, vaccine-induced antibody levels after the primary course provided only limited discrimination of SARS-CoV-2 infection risk during the Omicron wave.

This prospective cohort also serves as a use case to assess, from a methodological perspective, whether routinely collected data (such as those available via the Bordeaux University Hospital data warehouse) could have allowed to address the same scientific questions with a different research design.

8.3.3 Knowledge transfer

Participants: Rodolphe Thiébaud, Laura Richert, H el ene Savel.

We have set-up a transfert unit (BVA, Bordeaux Vaccine Analytics) with Adera, University of Bordeaux, to facilitate the collaborations with private companies. We have continued to develop a data warehouse system based on the Labkey solution where all raw data are organized and that includes meta-data on the design of the clinical trials and is used in international collaborations of facilitate data sharing and exploration (EHVA, EBOVAC, IP-Cure-B and CARE consortia).

9 Partnerships and cooperations

9.1 International initiatives

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

Participants: Boris Hejblum.

DESTRIER

Title: DEfining Surrogacy of early Transcriptomics foR vaccInE Response

Duration: 2022 -> 2024

Coordinator: Denis Agniel (dagniel@rand.org)

Partners:

- RAND Corporation ( tats-Unis)

Inria contact: Boris Hejblum

Summary: This project seeks to develop statistical methods to evaluate to which extent can transcriptomics be used to capture vaccine effects. Gene expression is central to protein production and largely determines cellular function: it is thus a promising biomarker for quickly measuring effects of vaccines. Validated transcriptomic signatures could thus be developed to dramatically speed up vaccine trials for emerging infectious diseases like Ebola or COVID-19. Such a technology could also be used for identifying good vaccine responders in health care workers that can be deployed in case of an epidemic emergency, or identify poor responders among vaccine recipient that would benefit from an additional booster dose. In this project, we set to develop novel statistical methods for assessing the surrogacy potential of transcriptomic data in vaccine research. We will first develop methods to quantify how much of the vaccine effect is mediated by gene expression, establishing if gene expression is suitable for capturing the vaccine's effect. We will develop model-free approaches to estimating this quantity which will remove many of the modeling assumptions typically used in high-dimensional mediation analysis. Second, we will develop methods to construct an optimal gene expression signature for capturing the vaccine effect, and we will develop methods to operationalize its use in future studies, establishing how to build and use such a transcriptomic signature. These methods will similarly take advantage of modern machine learning approaches and doubly robust estimation to provide model-free estimators of key quantities. We will use these methods to study high-impact clinical trials from the Vaccine Research Institute in the context of Ebola and COVID-19.

9.1.2 STIC/MATH/CLIMAT AmSud projects

Participants: Marta Avalos.

Title: Program MATH AmSud 2023, Chile, Uruguay, France (SMILE, code 23-MATH-12)

Partner Institution(s): Universidad de Valparaiso, and Universidad Adolfo Ibáñez, Chile,

Date/Duration: 2 years (until Dec 2025)

Additional info/keywords: Statistical Modeling, nonparametric Inference, and model sElection for complex data

9.1.3 Participation in other International Programs

RISE Rank-Based Identification of Surrogates in Small Ebola Studies (RISE)with Partner Institution(s) UT Austin, USA since June 2024. Award from Dr. Cécile DeWitt-Morette France-UT Endowed Excellence Fund with Layla Parast

Participants: Boris Hejblum.

MUSICC has been selected for funding by CEPI (Coalition for Epidemic Preparedness Innovations). This project, in which SISTM is a partner, will develop and conduct Controlled Human Infection Models For Beta-coronaviruses in order to assess vaccine effects. This project is rather unique in Europe by both the quality of the participants and its approach. In this context, SISTM will contribute to the data analysis and the modeling of the immune response. The project has started on February 1st, 2024. Duration 60 months, 01/02/2024 - 31/01/29. 355,000 USD.

Participants: Rodolphe Thiébaud, Mélanie Prague, Edouard Lhomme.

9.2 International research visitors

9.2.1 Visits of international scientists

Cristian Meza

Status Professor

Institution of origin: Universidad de Valparaiso

Country: Chile

Dates: 4th-8th March 2024

Context of the visit: MATH AmSud

Mobility program/type of mobility: research stay

Susana Eyheramendy

Status Professor

Institution of origin: Universidad Adolfo Ibáñez

Country: Chile

Dates: 4th-8th March 2024

Context of the visit: MATH AmSud

Mobility program/type of mobility: research stay

Lander Rodríguez

Status PhD student

Institution of origin: Basque Center for Applied Mathematics (BCAM)

Country: Spain

Dates: 1st Apr 2024 until 30th June 2024

Context of the visit: (Severo Ochoa Grant

Mobility program/type of mobility: research stay

John Barrera

Status PhD student

Institution of origin: Universidad de Valparaiso

Country: Chile

Dates: 18th-30th September 2024

Context of the visit: MATH AmSud

Mobility program/type of mobility: research stay

9.2.2 Visits to international teams**Research stays abroad****Arthur Hughes**

Visited institution: UT Austin (Layla Parast)

Country: USA

Dates: May-June 2024

Context of the visit: DESTRIER

Mobility program/type of mobility: research stay

Boris Hejblum**Visited institution:** UT Austin (Layla Parast)**Country:** USA**Dates:** June 2024**Context of the visit:** DESTRIER**Mobility program/type of mobility:** research stay**Marta Avalos****Visited institution:** Universidad de Valparaiso, and Universidad Adolfo Ibáñez**Country:** Chile**Dates:** 12th-27th August 2024**Context of the visit:** MATH AmSud**Mobility program/type of mobility:** research stay

Rodolphe Thiébaud is Adjunct professor, Department of Epidemiology, Biostatistics and Occupational Health, McGill University since 2023

9.3 European initiatives**9.3.1 H2020 projects**

SOLVE: The project funded by Horizon Europe has started on January 1st, 2024 to decipher the mechanisms of induction of long-lasting immunity through a comparison of vaccine platforms and to advance new vaccine concepts. In the project, SISTM is workpackage leader (WP7 Data Science) and will analyze the consortium's data to model the immune response of the 4 main different types of COVID19 vaccine platforms and some variants. SISTM will thus contribute to the comparison of these platforms and the discussion to present recommendations to stakeholders to support future epidemic preparation decision-making. Duration: 60 months 01/01/24 - 31/12/28. 563 330 Euros.

Participants: Rodolphe Thiébaud, Mélanie Prague, Linda Wittkop.

IP-CURE-B: Immune profiling to guide host-directed interventions to cure HBV infections. Co-ordinated by Inserm, the project includes a total of 13 Beneficiaries. In this project, SISTM will work on the analysis of data from the clinical intervention and the modelisation of the response to the treatment. L Wittkop. Duration: 60 months 01/01/20-31/12/25. 409,632 Euros.

Participants: Mélanie Prague, Linda Wittkop, Boris Hejblum.

9.3.2 Other european programs/initiatives

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. Thiébaud. Duration: 66 months. 01 /06 /2018 - 30 /11/2024. 351,274 Euros.

Participants: Mélanie Prague, Laura Richert, Boris Hejblum, Rodolphe Thiébaud, Quentin Clairon, Edouard Lhomme.

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. Thiébaud. Duration: 72 months. 01 /01 /2019 - 31 /12 /2024. 328,000 Euros.

Participants: Mélanie Prague, Laura Richert, Boris Hejblum, Rodolphe Thiébaud, Edouard Lhomme.

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 health threat. This major initiative aims at addressing two key objectives: the development of therapeutics to provide an emergency response towards the current COVID-19 pandemic and the development of therapeutics to address the current and/or future coronavirus outbreaks. To address both goals, the CARE consortium has carefully designed a comprehensive research and development (R&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.

Participants: Laura Richert, Boris Hejblum, Rodolphe Thiébaud, Edouard Lhomme.

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 was involved in statistical analysis of the tests data and will lead the WP3 which aims at modelling the epidemics. Duration: 01/05/2020 - 31/01/2024. 37,500 Euros.

Participants: Laura Richert, Rodolphe Thiébaud, Edouard Lhomme.

CoVICIS: The CoVICIS (EU-Africa Concerted Action on SAR-CoV-2 Virus Variant and Immunological Surveillance) program is proposing a global approach with a powerful state-of-the-art virologic and immunologic platforms coupled with large genomic surveillance studies and diverse cohorts in EU and SSA CoVICIS aims to contribute to the early identification of emerging VOC and address key unanswered questions regarding: i) the susceptibility to infection with VOC after a prior infection in the setting of a long-COVID or after vaccination with different vaccines, ii) the risk posed by VOC in immunocompromised patients, and iii) the modalities of infection and immune responses in children. CoVICIS is coordinated by the CHUV (Switzerland). SISTM is involved in WP7 Data Science and Analysis which aims to utilize cutting edge computational and statistical analysis method to obtain comprehensive assessment of immunogenicity and immune correlates of protection. Coordinated by the CHUV, CoVICIS counts 14 partners amongst which we can find Inserm, UNIMI, UNIGE, and 4 South-African partners. 11/2021-10/2024. Total budget: 10M€, SISTM budget: 110k€.

Participants: Linda Wittkop, Rodolphe Thiébaud.

9.4 National initiatives

Labex Vaccine Research Institute (VRI): Funded by the PIA under Laboratory of excellence initiative, VRI conducts research to accelerate the development of effective vaccines against HIV/AIDS and (re)-emerging infectious diseases. The SISTM team is leading the Data science division of the VRI. To this purpose, SISTM has established strong collaboration with immunologists. SISTM carries out biostatistical analysis of the data produced by the different other VRI teams together with a modelling approach of the immune response to the vaccines or other interventions. 2012-2025, Main partners: the VRI was established by the French National Agency for Research on AIDS and viral hepatitis (ANRS - France Recherche Nord & Sud Sida-HIV Hépatites) and the University of ParisEst Créteil (UPEC). The other partners of the VRI are CEA, Inserm, Pasteur Institute, the University of Bordeaux, the Baylor Institute for immunology research and the University of Strasbourg. Total budget: 75M€, SISTM budget: 1.85M€ (about 170k€ a year since 2012).

Participants: Mélanie Prague, Laura Richert, Boris Hejblum, Rodolphe Thiébaud, Edouard Lhomme, Quentin Clairon, Marta Avalos, Robin Genuer, Linda Wittkop.

Ecole Universitaire de Recherche “Digital Public Health” Funded under the PIA3 The Digital Public Health Graduate Program provides an interdisciplinary and international training from Master to Doctorate in epidemiology, biostatistics, computing and social sciences to explore the impact of digital public health on society. The whole program is directed by Rodolphe Thiébaud. The whole SISTM team is implicated in these activities. 2018-2028. Main partners: University of Bordeaux, Inserm, Inria, Sciences Po Bordeaux and University Bordeaux Montaigne. Total budget: 4.52 M€, SISTM budget: The budget is mostly dedicated to grants to students, running costs and indemnification of teachers.

Participants: Rodolphe Thiébaud.

PEPR Santé Numérique SMATCH: The PEPR SN SMATCH coordinated by Inria and co-coordinated by Sarah Zohar (HEKA) and Rodolphe Thiébaud (SISTM) is part of the France 2030 initiative to develop digital health in France. SMATCH objectives are to develop and apply statistical and AI-based methods with the ultimate goal of accelerating the development of medical interventions

(drugs and digital medical devices) during their evaluation in clinical trials based on the following assumptions:

1. The use of information generated in preclinical studies (animal studies, organoids, in silico studies) combined with adaptive designs should help the early phases of development;
2. The integration of multi-source data including real-world and in silico data should help to complete trials;
3. Specific adaptive designs should be defined for the evaluation of digital medical devices based on learning algorithms.

The consortium counts 16 teams mainly from Inria and Inserm Centers recognized in this field, bringing a unique and complementary expertise in data sciences and AI applied to health problems and specifically to clinical trials. In addition, links with the regulatory bodies involved are already established within the consortium (e.g. HAS) and outside (e.g. EMA). Finally, many connections exist with the other axes of the PEPR Digital Health and more generally with the projects carried out within the framework of the digital health acceleration strategy. Thus, by providing innovative and adapted methodological tools that will have already been applied in a real context, we hope to participate in the acceleration of clinical research leading to major societal and economic impacts. 01/09/2023 - 31/08/2029. Total budget : 3M€, SISTM budget: 693 996 €

Participants: Mélanie Prague, Laura Richert, Boris Hejblum, Rodolphe Thiébaud.

PIEEC MEDITWIN: MEDITWIN is a Projet Important d'Intérêt Européen Commun (PIEEC) part of the France 2030 strategy coordinated by Dassault Systems and Inria. The aim of the MEDITWIN project is to develop and validate digital twins to support personalised medical practices and strengthen the healthcare system in targeted therapeutic areas. These virtual twins will be multi-disciplinary and multi-physiological, and will be based on real clinical data, acquired prospectively and historically, at the molecular, genetic, cellular and tissue levels, right down to the organ, system, individual and population level. They will be based on structured, interoperable data hosted in sovereign infrastructures. In this frame, SISTM will develop innovative methods for adaptive clinical study designs for pilot (feasibility) and perpetual (after initial validation) clinical trial designs for the evaluation of patients' risk confronted to SaMD updates in collaboration with HEKA. 2024-2029, SISTM budget: 433 125 €

Participants: Mélanie Prague, Laura Richert, Rodolphe Thiébaud, Linda Witkop.

IHU VBHI: The Vascular Brain Health Institute (VBHI) is a joint-venture between the University of Bordeaux (UB), Bordeaux University Hospital (CHUB), the national institutes for medical and digital science research (Inserm, Inria), and the New Aquitaine region, aiming to create a Center of Excellence on Vascular Brain Health. It will establish an entirely novel paradigm to prevent stroke and dementia, two leading causes of death and disability worldwide, by taking a precision population health approach and leading an emerging global dynamic geared towards both innovation and inclusion. 11/2023-10/2032. The SISTM team will be involved mostly in WP1 to contribute to the analysis of high dimensional data and notably by conducting extensive bioinformatics analyses, including an original pipeline to identify miRNA-based candidate treatments for identified targets. In addition, the team will be involved in the design of omics- guided clinical trials design. Total budget: 40 M€ overall.

Participants: Rodolphe Thiébaud.

9.4.1 Various Partnership

Mélanie Prague: Chaire Digital Innovation and Health Data Science program of the Center for Applied Mathematics CMAP at the Ecole Polytechnique

The project team members are involved in:

- **F-CRIN** (French clinical research infrastructure network), initiated in 2012 by ANR under "Programme des Investissements d'avenir". (L Richert).
- **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 2020-2022, extended in 2023-2024. (Principal PI E. Lagarde Inserm U1219 in collaboration with University Hospital of Bordeaux. Marta Avalos is listed as a collaborator).
- **CESIR IV** (Combination of Studies on Health and Road Safety - 4th project) funded by ONISR DSR. (Principal PI E. Lagarde Inserm U1219. Marta Avalos is listed as a collaborator).
- **EMERG** (Exposome microbien et Risque sanitaire : intérêt d'une Gestion One Health des enjeux liés aux gripes zoonotiques) funded by PSGAR (Programmes Scientifiques de Grande Ambition Régionale). (Principal PI L Delhaes and D Malvy. Marta Avalos 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.
- **STRIVE** (Strategies and Treatments for Respiratory and Viral Emergencies Study Payments). International Network for respiratory and viral emergency studies. (Collaborator: Linda Wittkop).

10 Dissemination

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

Member of the organizing committees Robin Genuer served as the co-president of the organizing committee for the Journées de la Statistique 2024 (attended by more than 500 participants), held in Bordeaux in May 2024. Marta Avalos, Boris Hejblum, and Mélanie Prague were members of the organizing committee. Postdoctoral fellows and PhD students from the team also contributed to the organization.

Edouard Lhomme and Laura Richert organized the EUCLID annual scientific day, Bordeaux Nov 2024

10.1.2 Scientific events: selection

Member of the conference program committees Rodolphe Thiébaud is a member of the scientific committee of the IWHOD International Workshop on HIV Observational Databases since 2013.

Marta Avalos was a member of the program committee of

DATAQUITAINE Bordeaux, March 2023

FLAIRS-37 The 37th International Conference of the Florida Artificial Intelligence Research Society, AI in Healthcare Informatics track, Florida, May 2024

Cap2024 Conférence d'apprentissage automatique, Lille, Jul 2024

ML4H Machine Learning for Health, Vancouver, Canada, Dec 2024

Linda Wittkop was a member of the scientific committee of IWHOD in July 2024

Reviewer Marta Avalos was a reviewer for the conferences:

CHIL2024 Conference on Health, Inference, and Learning, New York, USA, Jun 2024

AAAI ReLM2024 Responsible Language Models, Feb 26, 2024, Vancouver, Canada

10.1.3 Journal

Member of the editorial boards Melanie Prague is an associate editor of International Journal of Biostatistics.

Reviewer - reviewing activities Boris Hejblum was a reviewer for the journals: BMC Genomics Data, Statistics In Medicine

Robin Genuer was a reviewer for the Journal of the American Statistical Association and Advances in Data Analysis and Classification

Quentin Clairon was a reviewer for AIMS Mathematics, Mathematical Bioscience and Engineering, NPJ Complexity and Journal of Intelligent & Fuzzy Systems

Laura Richert was a reviewer for Plos Med, Nature Communications, SMMR.

Marta Avalos was a reviewer for the 'Public Health and Epidemiology Informatics' section of the IMIA Yearbook.

Melanie Prague was a reviewer for Biometrics, CPT:pharmacology and pharmacometrics, biometrical journal, Frontiers in immunology, elife, and Plos computational Biology.

10.1.4 Invited talks

Boris Hejblum presented at ISNPS (International Symposium on Nonparametric Statistics) 2024 "Conditional independence testing by comparing empirical conditional cumulative distribution functions" [communication orale dans le cadre d'une session invitée].

Laura Richert gave an invited talk in Epiclin 2024 "Reducing the carbon footprint of clinical trials in France"

Marta Avalos presented at IMS International Conference on Statistics and Data Science (ICSDS), Nice, Dec 2024 [51] [communication orale dans le cadre d'une session invitée] and at the 1st SMILE Workshop on Statistical Modeling, Nonparametric Inference, and Model Selection for Complex Data, Valparaiso (Chile), Aug 2024 [50].

Melanie Prague was invited in a mini-symposium "Modeling within host" at Society of mathematical biology conference in Seoul 1-7 July 2024. She presented on "Integrating gene expression into mechanistic modeling of the immune system: Application to vaccination against COVID-19".

Rodolphe Thiébaud

- 13th GABRIEL International Meeting, November 27-29, 2024, Les Pensières Center for Global Health, Clinical study design in the context of emerging infectious diseases
- Conferencia virtual inteligencia artificial, salud pública y epidemiología, October 23, Lima, Artificial intelligence, Public Health and Epidemiology
- Small Data Symposium, October 7-8 2024, Freiburg, Integrating gene expression from whole blood into dynamical systems: illustration with a mechanistic model of the antibody response to COVID vaccination
- Journée FLI-CERF aux JFR 2024, 3 Octobre 2024, Paris, Le dépistage, vision transverse de Santé Publique
- Journée de l'Action coordonnée "Modélisation" ANRS/MIE, November 24th, 2023, Paris, How gene expression can inform dynamical models?

10.1.5 Leadership within the scientific community

Mélanie Prague co-leads with J. Guedj the working group Within-Host Modeling, Action Coordonnée Modélisation (ANRS-MIE) since 2021.

Hélène Savel is a board member of the “Biopharmacie et Santé” group of SFdS.

Auriane Gabaut is a board member of the “Jeunes Statisticien.ne.s” group of SFdS.

Laura Richert is co-coordinator of the “Vaccine working group” on viral hemorrhagic fevers (ANRS-MIE) and the working group “Greener clinical research” (Décarbonation de la recherche clinique) of the Recap/Inserm network”

Rodolphe Thiébaud is member of the Inserm Scientific committee, member of the Inserm ANRS-MIE scientific Advisory Board

10.1.6 Scientific expertise

Boris Hejblum is a national expert for the 2024 MESSIDORE project call from Inserm IReSP ‘Méthodologie des ESSais cliniques Innovants, Dispositifs, Outils et Recherches Exploitant les données de santé et biobanques’.

Boris Hejblum is a member of the ANRS MIE CSS13 (“Clinical research”) evaluation committee

Mélanie Prague is a member of the ANR committee CES45 Interfaces : mathématiques, sciences du numérique – biologie, santé

Rodolphe Thiébaud is a chair of the jury for the recruitment of permanent research at Inria Paris Saclay and a chair of the Inserm scientific committee for the evaluation of the Inserm Units in Paris

Linda Wittkop is a member of

- the CESREES Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé
- the external advisory board of Vaccelerate
- independent committees of clinical trials: ELDORADO and B-Free (President)

Marta Avalos was a member of the INRAE recruitment jury for two research engineer biostatisticians

10.1.7 Research administration

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

Mélanie Prague is a member of the bureau of Action Coordonnée Modélisation (ANRS-MIE) since 2021.

10.2 Teaching - Supervision - Juries

10.2.1 Teaching

Each faculty member is involved in teaching with approximatively MA 200 h/year, RG 200 h/year, RT 130 h/year, and LW 110 h/year, LR 80 h/year, BH 70 h/year, EL 80 h/year, MP 30 h/year. These activities splits as follow.

- **In class teaching**
 - Rodolphe Thiébaud is head of the Digital Public Health graduate program, University of Bordeaux. Robin Genuer is head of the M2 Biostatistique, Master of Public Health, University of Bordeaux, Linda Wittkop co-coordinates the M2 Epidemiology, Master of Public Health, University of Bordeaux
 - Master: All the permanent members and several PhD students teaches in the Master of Public Health (M1 Santé publique, M2 Biostatistique and/or M2 Epidemiology) and the Digital Public Health graduate program, University of Bordeaux.

- Master: Marta Avalos teaches in the Master of Applied Mathematics and Statistics (1st and/or 2nd year), University of Bordeaux.
 - Bachelor: Laura Richert, Linda Wittkop and Edouard Lhomme teach in PASS and DFASM1-3 (Diplôme de Formation Approfondie en Sciences Médicales) 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, Linda Wittkop, Rodolphe Thiébaud, Robin Genuer and Boris Hejblum coordinate several teaching units of Master in Public Health (Biostatistics, Epidemiology, Public Health Data-Science).
 - Laura Richert coordinates the teaching unit "critical article reading" (across 4 years of medical school), University of Bordeaux; Edouard Lhomme coordinates the teaching "Evaluation of health innovation (M2 Health innovations); Linda Wittkop coordinates the teaching unit "Public Health and Statistics in Medicine" of the first year of Medical School, University of Bordeaux; Marta Avalos co-coordinates the teaching unit "Statistical analysis of high-dimensional data" (M2 Statistical and stochastic modeling MSS and M2 in Statistical and Computer Engineering CMI ISI of the UFMI of the University of Bordeaux).
 - Boris Hejblum teaches a 4-day graduate course "Bayesian analysis for biomedical research" at the University of Copenhagen.
 - Boris Hejblum teaches on Statistical learning in high-dimension in M2 Numerical sciences & bio-health, École Centrale Nantes
 - Mélanie Prague teaches "Missing Data" at ENSAI Master Level and M2 Biostat at ISPED
 - ISPED Summer school: Introduction to R (Mélanie Prague), Data science with R (B Hejblum), Introduction to Rcmdr (Marta Avalos), Bayesian Methods for Biomedical Research (B Hejblum), Platform trial (Edouard Lhomme).
 - Quentin Clairon was the organiser and main teacher of the formation: "Inference of mechanistic models in Public Health" within the digital public health graduate program, University of Bordeaux.
- **E-learning**
 - Marta Avalos is head of the e-learning program of the Master of Public Health, 1st year, University of Bordeaux.
 - Mélanie Prague is responsible of a Module on "regression" in the Diplome Universitaire of statistics in epidemiology.
 - Master: Marta Avalos teaches in the e-learning program of the Master of Public Health (1st and 2nd year).
 - ODL University Course: Robin Genuer is head of the Diplôme universitaire "Méthodes statistiques en santé". Mélanie Prague and Robin Genuer teach in the Diplôme universitaire "Méthodes statistiques de régression en épidémiologie".
 - ODL University Course: Edouard Lhomme co-coordinates and teaches in the Diplôme universitaire "Recherche Clinique".
 - ODL University project: Robin Genuer participated to the IdEx Bordeaux University "Défi numérique" project [BeginR](#).

10.2.2 Supervision

Supervision of intern students:

- M Prague: Anne Andre Ruiz (ISPED Biostat M2); Tristan Candido da Silva (ISPED M1); Arthur Leroudier (ENSAE, M1); Jay Awale (Institute of technology India L3)

- B Hejblum: Mathéo Le Floch (50% M1 with Cécile Proust-Lima) and Alarig Vigneras (50% M1 with Robin Genuer)
- R Genuer supervised Louis Cazade's internship (Master's Year 1, Comparison of Variable Selection Methods Based on Random Forests), co-supervised Alarig Vigneras's internship with Boris Hejblum (Master's Year 1, Unsupervised Random Forests for Clustering), and co-supervised Alisha Dziarski's internship with Christophe Tzourio (Master's Year 2, A Comparison of Machine Learning Tools to Predict Depression Screen Outcomes: Advantages of a Balanced Training Approach).
- Q Clairon: Hugo Alves, M1 student
- L Richert: Michael Burtin, M2 Univ Claude Bernard Lyon, "Adaptation of Phase I Dose-Finding Clinical Trial Designs to the Context of Vaccinology."
- L Wittkop co-supervised with Jonathan Sterne: Suzie Wang (Master 2 PhDs, Bordeaux) and JiaJie Kang (MSc Epidemiology, University Bristol). She supervised Anderson N'Gattia (M2 Epi, ISPED)
- M Avalos: Ariel Guerra-Adames, M2 Machine Learning and Data Mining at Université Jean Monnet, St Etienne (50% with Emmanuel Lagarde), Chloé Renault, Research and innovation in Food for Health, Institut polytechnique Unilasalle, Beauvais (80% with Laurence Delhaes and Raphaël Enaud)

Supervision of PhD students:

- Boris Hejblum:
 - PhD defense: Benjamin Hivert (co-direction 50%): "Clustering et analyse différentielle de données d'expression génique", Sept 2024
 - Arthur Hughes (PhD co-direction 50%): "Approches par groupes de gènes pour le développement de vaccins : association, prédiction et marqueurs de substitution" co-directed with R Thiébaud, from Oct 2023.
 - Kalidou Ba (PhD co-direction 50%): "Reservoir computing for cellular composition prediction from longitudinal transcriptomics data in vaccine trials", co-directed with Xavier Hinaut (Inria Bordeaux), from Nov 2022.
 - Anness Pal (PhD co-direction 50%): "Modèles de mélange bayésien pour la déconvolution de proportions cellulaires à partir de données transcriptomiques en masse" co-directed with R Thiébaud, from Dec 2023.
 - Sara Fallet Pal (PhD co-direction 50%): "Analyse différentielle par groupes de gènes de données scRNA-seq issues d'échantillons multiples", co-directed with Pierre Neuvial (Institut Mathématiques de Toulouse, CNRS), from Oct 2024.
- Rodolphe Thiébaud:
 - PhD defense: 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 Thiébaud, Dec 2024.
 - PhD in progress: Thomas Ferté "Contribution of health data warehouses for clinical research and epidemiological surveillance in the context of Covid-19", co-directed by Rodolphe Thiébaud and Vianney Jouhet, from Sept 2022 AHU funding.
 - PhD in progress: Antonin Colajanni, Université de Bordeaux, from Oct 2023, co-direction with Patricia Thebault (Labri), RRI funding.
 - PhD in progress: Anness Pal, from Oct 2023, INRIA PEPR.
- Mélanie Prague:
 - Auriane GABAUT (50%) from Oct 2023. co-supervision with Cécile Proust-Lima "Utilisation de modèles à variables latentes pour construire des modèles mécanistes en grande dimension : application en immunologie pour le développement de vaccins."

- unofficial mentoring of Iris ganser (see R. Thiébaud directions)
- Adrien MITARD (50%) from Oct 2023. Co supervision with Jérémie Guedh IAME Inserm Paris. "Modélisation intra-hôte de la réponse vaccinale".
- Quentin Clairon:
 - Zhe Aurore Li (50%), first year PhD student. Mechanistic modeling using machine learning.
- Robin Genuer:
 - Justine REMIAT (1st year PhD started in 10/2023) "Development of distance-based random forests methods for complex data analysis", co-supervision with Cécile Proust-Lima
 - Corentin SEGALAS (Post-doc started in 02/2023), "Random forests for longitudinal data irregularly measured" co-supervision with Cécile Proust-Lima
- Laura Richert:
 - PhD in progress: Cyrille Kone "Bandit algorithms for early phase clinical trials in vaccinology", codirected by Emilie Kaufmann (Inria School) and Laura Richert from November 2022.
 - PhD in progress: Nam-Anh Tran (co-tutelle with McGill), 2024-2027, "Bayesian adaptive vaccine trials".
- Marta Avalos:
 - Ariel Guerra-Adames (1st year PhD started in Sept 2024) "Judgment biases in clinical decision-making during emergency triage and emergency medical dispatching: a quantitative and qualitative public health approach using large language models", co-supervision with Emmanuel Lagarde
- Linda Wittkop:
 - Adam Loffler, MD thesis "Utiliser un entrepôt de données de santé pour reproduire les résultats d'une cohorte vaccinale COVID-19 chez des patients atteints d'hypogammaglobulinémie".
- Edouard Lhomme :
 - Co supervised (50%) with Linda Wittkop Daniela Gouna "Modélisation et analyse des déterminants de l'immunogénicité et de la tolérance post-vaccination, à partir d'études cliniques vaccinales."

10.2.3 Juries

- Boris Hejblum
 - Jury PhD defense:
 - * Nicolas Enjalbert-Courrech, thesis examiner, IMT Toulouse, "Inférence post-sélection pour l'analyse des données transcriptomiques", Dec 2024
 - PhD Follow-up:
 - * Manel Rakez (2021 - 2025 ; directors Virginie Rondeau and Brice Amadeo — Université de Bordeaux), "Modélisation du dépistage du cancer du sein et son impact sur la réduction de la mortalité : Amélioration avec la modélisation conjointe et les réseaux de neurones profonds"
 - * Lucie Fontaine (2023 - 2026; director Frédérique Alexandre — Université de Bordeaux), "Modèles génératifs bio-inspirés pour l'IA"
 - * Catalina Gonzales-Gomez (2022-2025; director Manuel Rosa-calatrava — Université de Lyon 1), "Développement d'outils in silico pour le repositionnement pharmaceutique de molécules basés sur des données d'expression omics et la notion de connectivité"

* Gustavo Magana-Lopez (2023-2025; director Loïc Paulevé — Université de Bordeaux).
 “Synthèse et apprentissage d’ensembles de réseaux booléens prédictifs pour la reprogrammation cellulaire”

- Rodolphe Thiébaut was a member of PhD defense committees for Patrick Saux and Sophia Yazzourh.
- Robin Genuer was reviewer of the PhD thesis of Juliette Murriss (defended the 18/10/2024 at Univ. Paris Cité), and of the PhD thesis of Van Tuan Nguyen (defended the 12/12/2024 at Univ. Paris Cité).
- Laura Richert
 - PhD defense committees: Benjamin Duputel (Univ Paris Cité)
 - HDR defense committee: Nathalie De Castro (Univ Bordeaux)
 - PhD Follow-up: Vincent BOuteloup (Univ Bordeaux), Clément Massonaud (Paris Cité), Guillaume Mulier (Paris Cité)
- Melanie Prague :
 - was examiner of the PhD of Marion Naveau INRAE paris on "Procédures de sélection de variables en grande dimension dans les modèles non-linéaires à effets mixtes. Application en amélioration des plantes."
 - She is part of PhD follow-up of comittee of, Ilona Suhanda (Sorbonne université), Benjamin Glemain (inserm iPLesp Paris) and Maxime Beaulieu (Inserm IAME)
- Linda Wittkop
 - Jury PhD Follow-up: Sandrine DOMEQ "Analyse des trajectoires de soins : définition d'un cadre méthodologique et application au parcours des patients victimes d'un accident vasculaire cérébral" Supervisor Florence Saillour, Igor Sibon
- Marta Avalos
 - International PhD Follow-up: John Erick Barrera Perez, Doctorado En Matematica, from 2023, 'New developments in the estimation of statistical models for complex longitudinal data', University of Valparaiso (Chile).
 - Dylan Russon, “Vers des Urgences humaines et digitales : construction d’un outil de modélisation des flux et des processus pour l’exploration et la validation des initiatives innovantes”, Université de Bordeaux, from 2023
 - Academic tutor of PhD student S. Zaïd, Ecole doctorale Sociétés, Politique, Santé Publique de l’Université de Bordeaux- EDSP2.

10.3 Popularization

10.3.1 Specific official responsibilities in science outreach structures

M Avalos is a member of the Administrative Council (Research Division) of the competitiveness cluster ENTER (Digital Excellence in Service of Environmental and Responsible Transitions) and a member of the Labeling Committee

10.3.2 Participation in Live events

- B Hejblum and L Wittkop participated to the *Nuit des chercheurs 2024* by animating the Inserm exhibition “Des virus émergents et des épidémies”.
- B Hejblum participates yearly to the “Chiche! 1 Scientifique, 1 Classe” high-school program.

10.3.3 Others science outreach relevant activities

The Teamwork Art in Public Health: a Qualitative Study, presented in the ENLIGHT Teaching and Learning Conference 2024 - Innovation and Creativity in Higher Education, Oct 2024 [65]

Transmettre l'art de faire équipe à nos étudiants en Santé Publique : Retour d'expérience sur l'alignement pédagogique, presented in the ETES 2024 - Enseigner les Transitions Écologiques et Sociales dans le Supérieur, Jul 2024 [66]

Optimisation des services des urgences hospitalières à l'aide de l'IA : une opportunité incontournable et un risque de systématisation des disparités de santé, presented by Ariel Guerra-Adames [67] and Prédiction de l'épidémie de SARS-CoV-2 : Reservoir Computing appliqué à plus de 400 variables, presented by Thomas Ferté in Dataquitaine, March, 2024

Laura Richert presented a webinar "Développement durable et essais cliniques, Leem F-CRIN, CNCR" and presented at Colloque Bdx Ecologie et Santé "INNOVATIONS, RECHERCHE : toutes nos énergies au service de la TRANSFORMATION ÉCOLOGIQUE !"

Melanie Prague participated to NUMIN with an article on "Préparation aux épidémies : le pouvoir de la science des données". She has contributed to a paper in "Tangente Hors-série n°86. Nouveaux défis de la statistique" on "La médecine du futur".

Mélanie Prague and Rodolphe Thiébaud has publication in press: Sud-Ouest 06/02/2024 "Covid-19 une étude met en lumière les milliers de vies sauvées par les confinements et la vaccination", L'express 07/02/2024 "Covid-19 le confinement et les vaccins ont-ils été efficaces ce que révèle une étude", Le monde 08/02/2024 "Covid-19 sans la vaccination le nombre de morts aurait été le double en France", La Dépêche 09/02/2024 "Covid-19, les données prédisent 159000 décès supplémentaires une étude mesure l'impact du confinement et du vaccin en France.

11 Scientific production

11.1 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] M. Alexandre, R. Marlin, M. Prague, S. Coleon, N. Kahlaoui, S. Cardinaud, T. Naninck, B. Delache, M. Surenaud, M. Galhaut, N. Dereuddre-Bosquet, M. Cavarelli, P. Maisonnasse, M. Centlivre, C. Lacabaratz, A. Wiedemann, S. Zurawski, G. Zurawski, O. Schwartz, R. W. Sanders, R. Le Grand, Y. Levy and R. Thiébaud. 'Modelling the response to vaccine in non-human primates to define SARS-CoV-2 mechanistic correlates of protection'. In: *eLife* 11 (8th July 2022). DOI: [10.7554/eLife.75427](https://doi.org/10.7554/eLife.75427). URL: <https://hal.science/hal-04510052>.
- [3] M. Alexandre, M. Prague, C. Mclean, V. Bockstal, M. Douoguih and R. Thiébaud. 'Prediction of long-term humoral response induced by the two-dose heterologous Ad26.ZEBOV, MVA-BN-Filo vaccine against Ebola'. In: *NPJ vaccines* 8.174 (8th Nov. 2023). DOI: [10.1038/s41541-023-00767-y](https://doi.org/10.1038/s41541-023-00767-y). URL: <https://hal.science/hal-04226696>.
- [4] H. Barry, G. Mutua, H. Kibuuka, Z. Anywaine, S. Sirima, N. Meda, O. Anzala, S. Eholie, C. Bétard, L. Richert, C. Lacabaratz, M. J. McElrath, S. De Rosa, K. Cohen, G. Shukarev, C. Robinson, A. Gaddah, D. Heerwegh, V. Bockstal, K. Luhn, M. Leyssen, M. Douoguih and R. Thiébaud. 'Safety and immunogenicity of 2-dose heterologous Ad26.ZEBOV, MVA-BN-Filo Ebola vaccination in healthy and HIV-infected adults: A randomised, placebo-controlled Phase II clinical trial in Africa'. In: *PLoS Medicine* 18.10 (29th Oct. 2021), e1003813. DOI: [10.1371/journal.pmed.1003813](https://doi.org/10.1371/journal.pmed.1003813). URL: <https://hal.inria.fr/hal-03481531>.
- [5] D. Commenges, C. Alkhassim, R. Gottardo, B. P. Hejblum and R. Thiébaud. '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>.
- [6] D. Commenges and H. Jacqmin-Gadda. *Dynamical Biostatistical Models*. Chapman and Hall/CRC, 2015. URL: <https://hal.inria.fr/hal-01580149>.

- [7] A. Jarne, D. Commenges, M. Prague, Y. Levy and R. Thiébaud. ‘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>.
- [8] Y. Lévy, Y. Lévy, C. Moog, C. Moog, A. Wiedemann, A. Wiedemann, O. Launay, O. Launay, F. Candotti, F. Candotti et al. ‘Safety and immunogenicity of CD40.HIVRI.Env, a dendritic cell-based HIV vaccine, in healthy HIV-uninfected adults: a first-in-human randomized, placebo-controlled, dose-escalation study (ANRS VRI06)’. In: *EClinicalMedicine* 77.8 (Nov. 2024), p. 102845. DOI: [10.1016/j.eclinm.2024.102845](https://doi.org/10.1016/j.eclinm.2024.102845). URL: <https://hal.science/hal-04832417>.
- [9] P. Loubet, L. Wittkop, E. Tartour, B. Parfait, B. Barrou, J.-Y. Blay, M. Hourmant, M. Lachâtre, D.-A. Laplaud, M. Laville, B. Laviolle, J.-D. Lelièvre, J. Morel, S. Nguyen, J.-P. Spano, B. Terrier, A. Thiebaut, J.-F. Viillard, F. Vrtovsniak, X. De Lamballerie and O. Launay. ‘A French cohort for assessing COVID-19 vaccine responses in specific populations’. In: *Nature Medicine* 27.8 (12th July 2021), pp. 1319–1321. DOI: [10.1038/s41591-021-01435-1](https://doi.org/10.1038/s41591-021-01435-1). URL: <https://hal.sorbonne-universite.fr/hal-03290330>.
- [10] C. Pasin, F. Dufour, L. Villain, H. Zhang and R. Thiébaud. ‘Controlling IL-7 injections in HIV-infected patients’. In: *Bulletin of Mathematical Biology* (2018).
- [11] A. Pollard, O. Launay, J.-D. Lelièvre, C. Lacabaratz, S. Grande, N. Goldstein, C. Robinson, A. Gaddah, V. Bockstal, M. Leyssen et al. ‘Safety and immunogenicity of a two-dose heterologous Ad26.ZEBOV and MVA-BN-Filo Ebola vaccine regimen in adults in Europe (EBOVAC2): a randomised, observer-blind, participant-blind, placebo-controlled, phase 2 trial’. In: *The Lancet Infectious Diseases* (Nov. 2020). DOI: [10.1016/S1473-3099\(20\)30476-X](https://doi.org/10.1016/S1473-3099(20)30476-X). URL: <https://hal.inria.fr/hal-03142752>.
- [12] M. Prague. ‘Joint modeling of viral and humoral response in Non-human primates to define mechanistic correlates of protection for SARS-CoV-2’. In: Society for Mathematical Biology Annual Meeting - 2023. Columbus, United States, 16th July 2023. URL: <https://hal.science/hal-04416297>.
- [13] M. Prague, D. Commenges, J. M. Gran, B. Ledergerber, J. Young, H. Furrer and R. Thiébaud. ‘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>.
- [14] M. Prague and M. Lavielle. ‘SAMBA: a Novel Method for Fast Automatic Model Building in Nonlinear Mixed-Effects Models’. In: *CPT: Pharmacometrics and Systems Pharmacology* 11.2 (2022). DOI: [10.1002/psp4.12742](https://doi.org/10.1002/psp4.12742). URL: <https://hal.inria.fr/hal-03410025>.
- [15] A. Rechtién, 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. Thiébaud, 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>.

11.2 Publications of the year

International journals

- [16] M. Alexandre, M. Prague, E. Lhomme, J.-D. Lelièvre, L. Wittkop, L. Richert, Y. Lévy and R. Thiébaud. ‘Definition of Virological Endpoints Improving the Design of HIV Cure Strategies Using Analytical Antiretroviral Treatment Interruption’. In: *Clinical Infectious Diseases* 79.6 (31st May 2024), pp. 1447–1457. DOI: [10.1093/cid/ciae235](https://doi.org/10.1093/cid/ciae235). URL: <https://hal.science/hal-04843977> (cit. on p. 15).

- [17] J. Ambia, S. M. Ingle, K. McGinnis, N. Pantazis, M. J. Silverberg, L. Wittkop, K. Kusejko, H. Crane, A. van Sighem, M. Sarcletti, A. Cozzi-Lepri, P. Domingo, I. Jarrin, C. Wyen, M. Hessamfar, L. Zhang, M. Cavassini, J. Berenguer, T. R. Sterling, P. Reiss, S. Abgrall, M. J. Gill, A. Justice, J. a. C. Sterne and A. Trickey. 'Discrimination of the Veterans Aging Cohort Study Index 2.0 for Predicting Cause-specific Mortality Among Persons With HIV in Europe and North America'. In: *Open Forum Infectious Diseases* 11.7 (1st July 2024), ofae333. DOI: [10.1093/ofid/ofae333](https://doi.org/10.1093/ofid/ofae333). URL: <https://hal.science/hal-04728165>.
- [18] M. Ballif, D. Braun, A. Calmy, E. Bernasconi, M. Cavassini, F. Tissot, M. Stoeckle, P. Schmid, C. A. Fux, M. van der Valk, K. Brinkman, T. Mudrikova, F. Bonnet, O. Leleux, M. Saúde, D. Hirter, N. Schwab, A. Limacher, F. Rintelen, R. Kouyos, D. Haerry, S. C. Zambrano, M. Egloff, C. Akre, I. Peytremann-Bridevaux, A. Rauch, G. Wandeler and B. Surial. 'Booster-free anti-retroviral therapy for persons living with HIV and multidrug resistance (B-Free): protocol for a multicentre, multistage, randomised, controlled, non-inferiority trial'. In: *BMJ Open* 14.11 (21st Nov. 2024), e094912. DOI: [10.1136/bmjopen-2024-094912](https://doi.org/10.1136/bmjopen-2024-094912). URL: <https://hal.science/hal-04904029>.
- [19] H. Barry, E. Lhomme, M. Surenaud, M. Nouctara, C. Robinson, V. Bockstal, I. Valea, S. Somda, H. Tinto, N. Meda, B. Greenwood, R. Thiébaud and C. Lacabartz. 'Helminth exposure and immune response to the two-dose heterologous Ad26.ZEBOV, MVA-BN-Filo Ebola vaccine regimen'. In: *PLoS Neglected Tropical Diseases* 18.4 (11th Apr. 2024), e0011500. DOI: [10.1371/journal.pntd.0011500](https://doi.org/10.1371/journal.pntd.0011500). URL: <https://hal.science/hal-04611372> (cit. on p. 14).
- [20] J. Ben Farhat, M. Hessamfar, D. Neau, S. Farbos, E. Lazaro, P. Duffau, N. Rouanes, C. Cazanave, T. Pistone, P. Rispal, M.-A. Vandenhende, C. Krzyzanowsky, O. Leleux, L. Wittkop, F. Bonnet and D. Barger. 'Exposure to COVID-19 Pandemic-Related Stressors and Their Association With Distress, Psychological Growth and Drug Use in People With HIV in Nouvelle Aquitaine, France (ANRS CO3 AQUIVIH-NA Cohort-QuAliV-QuAliCOV Study)'. In: *AIDS and Behavior* (6th Jan. 2025). DOI: [10.1007/s10461-024-04588-5](https://doi.org/10.1007/s10461-024-04588-5). URL: <https://hal.science/hal-04871453>.
- [21] F. Bouvier, E. Peyrot, A. Balendran, C. Ségalas, I. Roberts, F. Petit and R. Porcher. 'Do machine learning methods lead to similar individualized treatment rules? A comparison study on real data'. In: *Statistics in Medicine* 43.11 (12th Mar. 2024), pp. 2043–2061. DOI: [10.1002/sim.10059](https://doi.org/10.1002/sim.10059). URL: <https://hal.science/hal-04503566>.
- [22] L. Capitaine, J. Bigot, R. Thiébaud and R. Genuer. 'Fréchet random forests for metric space valued regression with non Euclidean predictors'. In: *Journal of Machine Learning Research* 25.355 (2024), pp. 1–41. URL: <https://hal.science/hal-03066146> (cit. on pp. 6, 13).
- [23] M. Chalouni, P. Loubet, E. Lhomme, L. Ninove, B. Barrou, J. Y. Blay, M. Hourmant, J. de Seze, M. Laville, B. Laviolle, J. D. Lelièvre, J. Morel, S. N. Quoc, J. P. Spano, B. Terrier, A. Thiebaut, J. F. Viillard, F. Vrtovsnik, S. Circosta, A. Barquin, M. Gharib, E. Tartour, B. Parfait, R. Thiébaud, L. Meyer, X. de Lamballerie, O. Launay and L. Wittkop. 'Association between humoral serological markers levels and risk of SARS-CoV-2 infection after the primary COVID-19 vaccine course among ANRS0001S COV-POPART cohort participants'. In: *BMC Infectious Diseases* 24.1 (27th Sept. 2024), p. 1049. DOI: [10.1186/s12879-024-09861-5](https://doi.org/10.1186/s12879-024-09861-5). URL: <https://hal.science/hal-04724620> (cit. on p. 15).
- [24] F. Di Meglio, E. Lhomme, H. G. Ouedraogo, A. M. Barry, S. Doumbia, S. L. Faye, A. K. Mbawah, T. Sagna, M. Tounkara, R. Strauss, C. O. Doumbia, S. Diouf, K. Cissé, J. May, D. I. Puradiredja and D. Fusco. 'Variations in COVID-19 vaccine hesitancy over time: a serial cross-sectional study in five West African countries'. In: *BMJ Open* 14.11 (7th Nov. 2024), e083766. DOI: [10.1136/bmjopen-2023-083766](https://doi.org/10.1136/bmjopen-2023-083766). URL: <https://hal.science/hal-04802489>.
- [25] O. Dorémus, D. Russon, B. Contrand, A. Guerra-Adames, M. Avalos-Fernandez, C. Gil-Jardiné and E. Lagarde. 'Harnessing Moderate-Sized Language Models for Reliable Patient Data De-identification in Emergency Department Records: An Evaluation of Strategies and Performance'. In: *JMIR AI* (2024). DOI: [10.2196/57828](https://doi.org/10.2196/57828). URL: <https://inria.hal.science/hal-04489355>. In press.

- [26] I. Ganser, D. L. Buckeridge, J. Heffernan, M. Prague and R. Thiébaud. 'Estimating the population effectiveness of interventions against COVID-19 in France: A modelling study'. In: *Epidemics* 46 (1st Mar. 2024), p. 100744. DOI: [10.1016/j.epidem.2024.100744](https://doi.org/10.1016/j.epidem.2024.100744). URL: <https://hal.science/hal-04505758> (cit. on pp. 8, 11, 14).
- [27] A. Guerra Adames, M. Avalos, O. Dorémus, C. Gil-Jardiné and E. Lagarde. 'Uncovering Judgment Biases in Emergency Triage: A Public Health Approach Based on Large Language Models'. In: *Proceedings of Machine Learning Research* (2024). URL: <https://hal.science/hal-04846735>. In press (cit. on pp. 11, 14).
- [28] B. Hejblum, K. Ba, R. Thiébaud and D. Agniel. 'Neglecting normalization impact in semi-synthetic RNA-seq data simulation generates artificial false positives'. In: *Genome Biology* 25.1 (30th Oct. 2024), p. 281. DOI: [10.1101/2022.05.10.490529](https://doi.org/10.1101/2022.05.10.490529). URL: <https://hal.science/hal-03908246>.
- [29] M. Hentzien, F. Bonnet, E. Bernasconi, E. Biver, D. L. Braun, A. Munting, K. Leuzinger, O. Leleux, S. Musardo, V. Prendki, P. Schmid, C. Staehelin, M. Stoeckle, C. S. Walti, L. Wittkop, V. Appay, A. M. Didierlaurent and A. Calmy. 'Immune response to the recombinant herpes zoster vaccine in people living with HIV over 50 years of age compared to non-HIV age-/gender-matched controls (SHINGR'HIV): a multicenter, international, non-randomized clinical trial study protocol'. In: *BMC Infectious Diseases* 24.1 (2024), p. 329. DOI: [10.1186/s12879-024-09192-5](https://doi.org/10.1186/s12879-024-09192-5). URL: <https://hal.science/hal-04674638>.
- [30] B. Hivert, D. Agniel, R. Thiébaud and B. P. Hejblum. 'Post-clustering difference testing: valid inference and practical considerations'. In: *Computational Statistics and Data Analysis* 193 (2024), p. 107916. DOI: [10.1016/j.csda.2023.107916](https://doi.org/10.1016/j.csda.2023.107916). URL: <https://hal.science/hal-03889565> (cit. on p. 13).
- [31] T. Klopfenstein, S. Zayet, S. Poloni, V. Gendrin, D. Fournier, J.-B. Vuilleminot, P. Selles, A. Dussaucy, G. Coureau, M. F. Avalos, L. Toko, P.-Y. Royer, C.-E. Lavoignet, B. Amari, M. Puyraveau and C. Chirouze. 'CRP under 130 mg/L rules out the diagnosis of Legionella pneumophila serogroup 1 (URINELLA Study)'. In: *European Journal of Clinical Microbiology and Infectious Diseases* (2024). DOI: [10.1007/s10096-024-04814-x](https://doi.org/10.1007/s10096-024-04814-x). URL: <https://hal.science/hal-04522802>. In press.
- [32] C. Lacabaratz, M. Durand, A. Wiedemann, E. Foucat, M. Surenaud, C. Krief, L. Guillaumat, C. Robinson, K. Luhn, V. Bockstal, R. Thiebaut, L. Richert and Y. Levy. 'Innate and cellular immune response to the Ebola vaccine Ad26.ZEBOV, MVA-BN-Filo: an ancillary study of the EBL2001 phase II trial'. In: *Journal of Infectious Diseases* (16th July 2024), jiae360. DOI: [10.1093/infdis/jiae360](https://doi.org/10.1093/infdis/jiae360). URL: <https://hal.science/hal-04734724> (cit. on p. 14).
- [33] Y. Lévy, Y. Lévy, C. Moog, C. Moog, A. Wiedemann, A. Wiedemann, O. Launay, O. Launay, F. Candotti, F. Candotti et al. 'Safety and immunogenicity of CD40.HIVRI.Env, a dendritic cell-based HIV vaccine, in healthy HIV-uninfected adults: a first-in-human randomized, placebo-controlled, dose-escalation study (ANRS VRI06)'. In: *EclinicalMedicine* 77.8 (Nov. 2024), p. 102845. DOI: [10.1016/j.eclinm.2024.102845](https://doi.org/10.1016/j.eclinm.2024.102845). URL: <https://hal.science/hal-04832417> (cit. on pp. 11, 15).
- [34] P. Loubet, J.-D. Lelievre, A. Francois, E. Botelho-Nevers, C. Chidiac, D. Chirio, V. Dubee, B. Dussol, F. Galtier, M. Hessamfar, E. Hodaj, S. Jaffuel, K. Lacombe, F. Laine, M. Lefebvre, Z. Maakaroun-Vermesse, A. Makinson, A. Portefaix, V. Pourcher, D. Rey, D. Zucman, J. Longobardi, M. Bertheau, E. Tartour, X. de Lamballerie, O. Launay and L. Wittkop. 'Humoral response after mRNA COVID-19 primary vaccination and single booster dose in people living with HIV compared to controls: a French nationwide multicenter cohort study - ANRS0001s COV-POPART'. In: *International Journal of Infectious Diseases* 146 (31st May 2024), p. 107110. DOI: [10.1016/j.ijid.2024.107110](https://doi.org/10.1016/j.ijid.2024.107110). URL: <https://hal.science/hal-04627501> (cit. on p. 15).
- [35] A. M. Manyara, P. Davies, D. Stewart, C. Weir, A. Young, J. Blazeby, N. Butcher, S. Bujkiewicz, A.-W. Chan, D. Dawoud, M. Offringa, M. Ouwens, A. Hróbjartsson, A. Amstutz, L. Bertolaccini, V. D. Bruno, D. Devane, C. Faria, P. Gilbert, R. Harris, M. Lassere, L. Marinelli, S. Markham, J. Powers, Y. Rezaei, L. Richert, F. Schwendicke, L. Tereshchenko, A. Thoma, A. Turan, A. Worrall, R. Christensen, G. Collins, J. Ross, R. Taylor and O. Ciani. 'Reporting of surrogate endpoints in

- randomised controlled trial protocols (SPIRIT-Surrogate): extension checklist with explanation and elaboration'. In: *BMJ - British Medical Journal* 386.1 (9th July 2024), e078525. DOI: [10.1136/bmj-2023-078525](https://doi.org/10.1136/bmj-2023-078525). URL: <https://hal.science/hal-04832415>.
- [36] A. M. Manyara, P. Davies, D. Stewart, C. J. Weir, A. E. Young, J. Blazeby, N. J. Butcher, S. Bujkiewicz, A.-W. Chan, D. Dawoud, M. Offringa, M. Ouwens, A. Hrobjartsson, A. Amstutz, L. Bertolaccini, V. D. Bruno, D. Devane, C. D. C. M. Faria, P. B. Gilbert, R. Harris, M. Lassere, L. Marinelli, S. Markham, J. H. Powers, Y. Rezaei, L. Richert, F. Schwendicke, L. G. Tereshchenko, A. Thoma, A. Turan, A. Worrall, R. Christensen, G. S. Collins, J. S. Ross, R. S. Taylor and O. Ciani. 'Reporting of surrogate endpoints in randomised controlled trial reports (CONSORT-Surrogate): extension checklist with explanation and elaboration'. In: *BMJ* 386 (9th July 2024), e078524. DOI: [10.1136/bmj-2023-078524](https://doi.org/10.1136/bmj-2023-078524). URL: <https://hal.science/hal-04738314>.
- [37] M. Nunes, E. Thommes, H. Fröhlich, A. Flahault, J. Arino, M. Baguelin, M. Biggerstaff, G. Bizel-Bizellot, R. Borchering, G. Cacciapaglia, S. Cauchemez, A. Barbier-Chebbah, C. Claussen, C. Choirat, M. Cojocaru, C. Commaille-Chapus, C. Hon, J. Kong, N. Lambert, K. Lauer, T. Lehr, C. Mahe, V. Marechal, A. Mebarki, S. Moghadas, R. Niehus, L. Opatowski, F. Parino, G. Pruvost, A. Schuppert, R. Thiébaud, A. Thomas-Bachli, C. Viboud, J. Wu, P. Crepey and L. Coudeville. 'Redefining pandemic preparedness: Multidisciplinary insights from the CERP modelling workshop in infectious diseases, workshop report'. In: *Infectious Disease Modelling* 9.2 (June 2024), pp. 501–518. DOI: [10.1016/j.idm.2024.02.008](https://doi.org/10.1016/j.idm.2024.02.008). URL: <https://ehesp.hal.science/hal-04551033>.
- [38] Y. Pastor, O. Reynard, M. Iampietro, M. Surenaud, F. Picard, N. El Jahrani, C. Lefebvre, A. Ham-moudi, L. Dupaty, E. Brisebard, S. Reynard, E. Moureaux, M. Moroso, S. Durand, C. Gonzalez, L. Amurri, A.-S. Gallouet, R. Marlin, S. Baize, E. Chevillard, H. Raoul, H. Hocini, M. Centlivre, R. Thiébaud, B. Horvat, V. Godot, Y. Levy and S. Cardinaud. 'A vaccine targeting antigen-presenting cells through CD40 induces protective immunity against Nipah disease'. In: *Cell Reports Medicine* 5.3 (19th Mar. 2024), p. 101467. DOI: [10.1016/j.xcrm.2024.101467](https://doi.org/10.1016/j.xcrm.2024.101467). URL: <https://hal.science/hal-04569963>.
- [39] M.-B. Requena, C. Protopopescu, A. C. Stewart, D. K. van Santen, M. B. Klein, I. Jarrin, J. Berenguer, L. Wittkop, D. Salmon, A. Rauch, M. Prins, M. van der Valk, R. Sacks-Davis, M. E. Hellard, P. Carrieri and K. Lacombe. 'All-cause mortality before and after DAA availability among people living with HIV and HCV: An international comparison between 2010 and 2019'. In: *International Journal of Drug Policy* 124 (6th Jan. 2024), p. 104311. DOI: [10.1016/j.drugpo.2023.104311](https://doi.org/10.1016/j.drugpo.2023.104311). URL: <https://hal.science/hal-04467710>.
- [40] R. Sacks-Davis, D. K. van Santen, A. Boyd, J. Young, A. Stewart, J. S. Doyle, A. Rauch, C. Mugglin, M. Klein, M. van der Valk, C. Smit, I. Jarrin, J. Berenguer, K. Lacombe, M.-B. Requena, L. Wittkop, O. Leleux, F. Bonnet, D. Salmon, G. V. Matthews, R. Guy, N. K. Martin, T. Spelman, M. Prins, M. Stooze and M. Hellard. 'Changes in incidence of hepatitis C virus reinfection and access to direct-acting antiviral therapies in people with HIV from six countries, 2010-19: an analysis of data from a consortium of prospective cohort studies'. In: *Lancet HIV* 11.2 (12th Jan. 2024), e106–e116. DOI: [10.1016/s2352-3018\(23\)00267-9](https://doi.org/10.1016/s2352-3018(23)00267-9). URL: <https://hal.science/hal-04468540>.
- [41] D. K. van Santen, A. Stewart, J. S. Doyle, M. A. Stooze, J. Asselin, M. B. Klein, J. Young, J. Berenguer, I. Jarrin, K. Lacombe, L. Wittkop, O. Leleux, D. Salmon, F. Bonnet, A. Rauch, C. Mugglin, G. Matthews, M. Prins, C. Smit, A. Boyd, M. van der Valk, R. Sacks-Davis and M. E. Hellard. 'Cohort Profile: International Collaboration on Hepatitis C Elimination in HIV Cohorts (InCHEHC)'. In: *International Journal of Epidemiology* 53.1 (1st Feb. 2024). DOI: [10.1093/ije/dyad154](https://doi.org/10.1093/ije/dyad154). URL: <https://hal.science/hal-04515032>.
- [42] H. Savel, F. Meyer-Losic, C. Proust-Lima and L. Richert. 'Statistical classification of treatment responses in mouse clinical trials for stratified medicine in oncology drug discovery'. In: *Scientific Reports* 14.1 (9th Jan. 2024), p. 934. DOI: [10.1038/s41598-023-51055-7](https://doi.org/10.1038/s41598-023-51055-7). URL: <https://inria.hal.science/hal-04403059>.
- [43] C. Ségalas, C. Helmer, R. Genuer and C. Proust-Lima. 'Functional Principal Component Analysis as an Alternative to Mixed-Effect Models for Describing Sparse Repeated Measures in Presence of Missing Data'. In: *Statistics in Medicine* (9th Sept. 2024). DOI: [10.1002/sim.10214](https://doi.org/10.1002/sim.10214). URL: <https://hal.science/hal-04693072> (cit. on p. 13).

- [44] R. Thiébaud, B. P. Hejblum, F. Mougin, C. Tzourio and L. Richert. 'ChatGPT and beyond with artificial intelligence (AI) in health: Lessons to be learned'. In: *Revue du Rhumatisme* 91.1 (1st Jan. 2024), pp. 12–15. DOI: [10.1016/j.rhum.2023.07.007](https://doi.org/10.1016/j.rhum.2023.07.007). URL: <https://hal.science/hal-04466670> (cit. on p. 14).
- [45] A. Trickey, R. Glaubius, N. Pantazis, R. Zangerle, L. Wittkop, J. Vehreschild, S. Grabar, M. Cavasini, R. Teira, A. d'Arminio Monforte, J. Casabona, A. van Sighem, I. Jarrin, S. M. Ingle, J. a. C. Sterne, J. W. Imai-Eaton and L. F. Johnson. 'Estimation of Improvements in Mortality in Spectrum Among Adults With HIV Receiving Antiretroviral Therapy in High-Income Countries'. In: *Journal of Acquired Immune Deficiency Syndromes - JAIDS* 95.1S (1st Jan. 2024), e89–e96. DOI: [10.1097/qai.0000000000003326](https://doi.org/10.1097/qai.0000000000003326). URL: <https://hal.science/hal-04467807>.
- [46] S. Valayer, M. Alexandre, M. Prague, A. H. Beavogui, S. Doumbia, M. Kieh, B. Greenwood, B. Leigh, M. Poupelin, C. Schwimmer, S. Sow, I. M. Berry, J. Kuhn, D. Fusco, N. D. Cauwelaert, D. Watson-Jones, R. Thiébaud, Y. Lévy, Y. Yazdanpanah, L. Richert and E. Lhomme. 'Evaluation of waning of IgG antibody responses after rVSVΔG-ZEBOV-GP and Ad26.ZEBOV, MVA-BN-Filo Ebola virus disease vaccines: a modelling study from the PREVAC randomized trial'. In: *Emerging microbes & infections* (19th Nov. 2024), Online ahead of print. DOI: [10.1080/22221751.2024.2432353](https://doi.org/10.1080/22221751.2024.2432353). URL: <https://inria.hal.science/hal-04817450> (cit. on p. 14).
- [47] L. Weskamm, P. Tarnow, C. Harms, M. Huchon, M. Raadsen, M. Friedrich, L. Rübenacker, C. Grüttner, M. Garcia, T. Koch, S. Becker, G. Sutter, E. Lhomme, B. Haagmans, A. Fathi, S. Blois, C. Dahlke, L. Richert and M. Addo. 'Dissecting humoral immune responses to an MVA-vectored MERS-CoV vaccine in humans using a systems serology approach'. In: *iScience* 27.8 (Aug. 2024), p. 110470. DOI: [10.1016/j.isci.2024.110470](https://doi.org/10.1016/j.isci.2024.110470). URL: <https://inria.hal.science/hal-04707178>.
- [48] E. White, L. Papagno, A. Samri, K. Sugata, B. P. Hejblum, A. R. Henry, D. C. Rogan, S. Darko, P. Recordon-Pinson, Y. Dudoit, S. Llewellyn-Lacey, L. A. Chakrabarti, F. Buseyne, S. A. Migueles, D. A. Price, M.-A. Andreola, Y. Satou, R. Thiébaud, C. Katlama, B. Autran, D. C. Douek and V. Appay. 'Clonal succession after prolonged antiretroviral therapy rejuvenates CD8⁺ T cell responses against HIV-1'. In: *Nature Immunology* 25.9 (1st Sept. 2024), pp. 1555–1564. DOI: [10.1038/s41590-024-01931-9](https://doi.org/10.1038/s41590-024-01931-9). URL: <https://hal.science/hal-04703528>.
- [49] A. Wiedemann, E. Lhomme, M. Huchon, E. Foucat, M. Bérerd-Camara, L. Guillaumat, M. Yaradouno, J. Tambalou, C. Rodrigues, A. Ribeiro et al. 'Long-term cellular immunity of vaccines for Zaire Ebola Virus Diseases'. In: *Nature Communications* 15.1 (3rd Sept. 2024), p. 7666. DOI: [10.1038/s41467-024-51453-z](https://doi.org/10.1038/s41467-024-51453-z). URL: <https://inria.hal.science/hal-04707176> (cit. on p. 14).

Invited conferences

- [50] M. Avalos. 'A Biostatistician in the Era of a Paradigm Shift Towards Data Science in Epidemiology'. In: 1st SMILE Workshop on Statistical Modeling, Nonparametric Inference, and Model Selection for Complex Data. Valparaiso, Chile, Aug. 2024. URL: <https://inria.hal.science/hal-04880198> (cit. on p. 24).
- [51] M. Avalos, J. Barrera, C. Meza, S. Eyheramendy, L.-E. Vandenberght, R. Enaud and L. Delhaes. 'Challenges of Microbiome Data Analysis in Chronic Respiratory Disease Studies: Examples from Three French Studies'. In: International Conference on Statistics and Data Science (ICSIDS). Nice, France, 16th Dec. 2024. URL: <https://hal.science/hal-04846596> (cit. on pp. 14, 24).

International peer-reviewed conferences

- [52] M. Avalos, D. Cohen, D. Russon, M. Davids, O. Dorémus, G. Chenais, E. Tellier, C. Gil-Jardiné and E. Lagarde. 'Detecting Human Bias in Emergency Triage Using LLMs: Literature Review, Preliminary Study, and Experimental Plan'. In: FLAIRS 2024 - 37th International Florida Artificial Intelligence Research Society Conference. Vol. 37. Miramar Beach, United States: LibraryPress@UF; FLVC Library Services, May 2024, p. 6. URL: <https://inria.hal.science/hal-04575557>.

- [53] T. Ferté, D. Dutartre, B. P. Hejblum, R. Griffier, V. Jouhet, R. Thiébaud, X. Hinaut and P. Legrand. ‘Optimizing Reservoir Computing with Genetic Algorithm for High-Dimensional SARS-CoV-2 Hospitalization Forecasting: Impacts of Genetic Algorithm Hyperparameters on Feature Selection and Reservoir Computing Hyperparameter Tuning’. In: 16th International Conference, Évolution Artificielle, EA 2024. Lecture Notes in Computer Science. Bordeaux, France, 29th Oct. 2024. URL: <https://inria.hal.science/hal-04905975>.
- [54] A. Guerra-Adames, M. Avalos, D. Cohen, D. Russon, M. Davids, O. Doremus, G. Chenais, E. Tellier, C. Gil-Jardiné and E. Lagarde. ‘Reflections for the design of an experimental protocol for Bias Detection in Hospital Emergency Triage using Language Models’. In: EvalLLM2024 - Atelier sur l'évaluation des modèles génératifs (LLM) et challenge d'extraction d'information few-shot. Toulouse, France, 8th July 2024. URL: <https://hal.science/hal-04678043>.
- [55] C. Kone, E. Kaufmann and L. Richert. ‘Bandit Pareto Set Identification: the Fixed Budget Setting’. In: *Proceedings of Machine Learning Research*. AISTATS. Valencia (Espagne), Spain, 2nd May 2024. URL: <https://hal.science/hal-04877368>.

Conferences without proceedings

- [56] T. Ferté, D. Dutartre, B. P. Hejblum, R. Griffier, V. Jouhet, R. Thiébaud, P. Legrand and X. Hinaut. ‘Reservoir Computing for Short High-Dimensional Time Series: an Application to SARS-CoV-2 Hospitalization Forecast’. In: ICML'24: Proceedings of the 41st International Conference on Machine Learning. Vol. 235. Proceedings of Machine Learning Research. Vienna, Austria, France, 8th July 2024, pp. 13570–13591. DOI: [10.5555/3692070](https://doi.org/10.5555/3692070). URL: <https://hal.science/hal-04693930> (cit. on pp. 13, 14).

Doctoral dissertations and habilitation theses

- [57] I. Ganser. ‘INSIGHTS FROM MATHEMATICAL MODELS INTO COVID-19: ANALYZING PUBLIC HEALTH INTERVENTIONS AND IMMUNITY DYNAMICS’. Université de Bordeaux; McGill University, 11th Dec. 2024. URL: <https://theses.hal.science/tel-04904849>.
- [58] B. P. Hejblum. ‘Statistical methods for leveraging high-dimensional data from high-throughput measurements in vaccine clinical development’. Université de Bordeaux, 7th May 2024. URL: <https://hal.science/tel-04633105> (cit. on p. 12).
- [59] B. Hivert. ‘Clustering and differential analysis of gene expression data’. Université de Bordeaux, 24th Sept. 2024. URL: <https://theses.hal.science/tel-04767353>.
- [60] M. Prague. ‘Mechanistic and Statistical Models for Treatment and Control of Infectious Diseases’. Université de Bordeaux, 5th June 2024. URL: <https://hal.science/tel-04673733> (cit. on p. 11).

Reports & preprints

- [61] B. Hivert, D. Agniel, R. Thiébaud and B. P. Hejblum. *Running in circles: practical limitations for real-life application of data fission and data thinning in post-clustering differential analysis*. 20th Oct. 2024. URL: <https://hal.science/hal-04745517> (cit. on p. 13).

Other scientific publications

- [62] A. Guerra-Adames, M. Avalos, O. Dorémus, C. Gil-Jardiné and E. Lagarde. ‘Uncovering Judgment Biases in Emergency Triage: A Public Health Approach Based on Large Language Models’. In: ML4H 2024 - Machine Learning for Health Symposium. Vancouver, Canada, 15th Dec. 2024. URL: <https://hal.science/hal-04846696>.
- [63] D. Russon, M. Avalos, A. Guerra-Adames, C. Gil-Jardiné and E. Lagarde. ‘AI-Driven Emergency Patient Flow Optimization is Both an Unmissable Opportunity and a Risk of Systematizing Health Disparities’. In: FLAIRS-37 2024 - 37th International FLAIRS Conference. Vol. 37. Miramar Beach, United States: LibraryPress@UF; FLVC Library Services, May 2024, p. 4. DOI: [10.32473/flairs.37](https://doi.org/10.32473/flairs.37). URL: <https://inria.hal.science/hal-04575571>.

- [64] L.-E. Vandenberght, M. Avalos, C. Labarca, V. Elankumaran, R. Enaud and L. Delhaes. ‘Exploring Interaction Networks of Indoor Environmental Microbiome and Mycobiome in Asthmatic and Healthy Subjects: Insights from the COBRA-ENV 2 Study’. In: IUMS 2024 - International Union of Microbiological Societies. Firenze, Italy, 23rd Oct. 2024. URL: <https://inria.hal.science/hal-04755097>.

Scientific popularization

- [65] P. Arnal, C. Cochet, A. Fergelot, L. Sébastien, A. Rivadeneyra and M. Avalos. ‘The Teamwork Art in Public Health: a Qualitative Study.: Fostering Collaboration and Innovation Through Academic Group Projects.’ In: The 4th ENLIGHT Teaching and Learning Conference at TARTU 2024. Tartu, Estonia, 7th Oct. 2024. URL: <https://hal.science/hal-04731734> (cit. on p. 30).
- [66] M. Avalos, V. Kiewsky, M. Plazy, E. Verpillot and E. Hierholzer. ‘Transmettre l’art de faire équipe à nos étudiants en Santé Publique : Retour d’expérience sur l’alignement pédagogique’. In: ETES 2024 - Enseigner les Transitions Écologiques et Sociales dans le Supérieur. Pessac, France, 9th July 2024. URL: <https://hal.science/hal-04644137> (cit. on p. 30).
- [67] A. Guerra-Adames, M. Avalos and E. Lagarde. ‘Optimisation des services des urgences hospitalières à l’aide de l’IA : une opportunité incontournable et un risque de systématisation des disparités de santé’. In: Dataquitaine 2024. Talence (33), France, Mar. 2024. URL: <https://inria.hal.science/hal-04546590> (cit. on p. 30).

11.3 Cited publications

- [68] O. O. Aalen, K. Røysland, J. M. Gran and B. Ledergerber. ‘Causality, mediation and time: a dynamic viewpoint’. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 175.4 (2012), pp. 831–861. DOI: <https://doi.org/10.1111/j.1467-985X.2011.01030.x> (cit. on p. 7).
- [69] D. Agniel, B. Hejblum, R. Thiébaud and L. Parast. ‘Doubly-robust evaluation of high-dimensional surrogate markers’. In: *Biostatistics* 24.4 (2023), pp. 985–999. DOI: [10.1093/biostatistics/kxac020](https://doi.org/10.1093/biostatistics/kxac020). URL: <https://inria.hal.science/hal-03100499> (cit. on p. 6).
- [70] M. Alexandre, R. Marlin, M. Prague, S. Coleon, N. Kahlaoui, S. Cardinaud, T. Naninck, B. Delache, M. Surenaud, M. Galhaut et al. ‘Modelling the response to vaccine in non-human primates to define SARS-CoV-2 mechanistic correlates of protection’. In: *elife* 11 (2022), e75427 (cit. on p. 8).
- [71] M. Alexandre, M. Prague, C. McLean, V. Bockstal, M. Douoguih, R. Thiébaud, E. 1 and E. 2. Consortia. ‘Prediction of long-term humoral response induced by the two-dose heterologous Ad26. ZEBOV, MVA-BN-Filo vaccine against Ebola’. In: *NPJ vaccines* 8.1 (2023), p. 174 (cit. on p. 8).
- [72] C. Bouveyron. ‘Adaptive Mixture Discriminant Analysis for Supervised Learning with Unobserved Classes’. In: *Journal of Classification* 31.1 (2014), pp. 49–84. DOI: [10.1007/s00357-014-9147-x](https://doi.org/10.1007/s00357-014-9147-x) (cit. on p. 6).
- [73] Q. Clairon, M. Prague, D. Planas, T. Bruel, L. Hocqueloux, T. Prazuck, O. Schwartz, R. Thiébaud and J. Guedj. ‘Modeling the kinetics of the neutralizing antibody response against SARS-CoV-2 variants after several administrations of Bnt162b2’. In: *PLoS Computational Biology* 19.8 (2023), e1011282 (cit. on p. 8).
- [74] C. Colas, B. Hejblum, S. Rouillon, R. Thiébaud, P.-Y. Oudeyer, C. Moulin-Frier and M. Prague. ‘Epidemioptim: A toolbox for the optimization of control policies in epidemiological models’. In: *Journal of Artificial Intelligence Research* 71 (2021), pp. 479–519 (cit. on p. 8).
- [75] A. Collin, B. P. Hejblum, C. Vignals, L. Lehot, R. Thiébaud, P. Moireau and M. Prague. ‘Using population based Kalman estimator to model COVID-19 epidemic in France: estimating the effects of non-pharmaceutical interventions on the dynamics of epidemic’. In: *The international journal of biostatistics* (2023) (cit. on p. 8).
- [76] T. Ferté, S. Cossin, T. Schaefferbeke, T. Barnetche, V. Jouhet and B. Hejblum. ‘Automatic phenotyping of electronic health record: PheVis algorithm’. In: *Journal of Biomedical Informatics* 117 (2021), p. 103746. DOI: [10.1016/j.jbi.2021.103746](https://doi.org/10.1016/j.jbi.2021.103746). URL: <https://hal.inria.fr/hal-03100435> (cit. on p. 6).

- [77] T. Ferté, V. Jouhet, R. Griffier, B. Hejblum and R. Thiébaud. ‘The benefit of augmenting open data with clinical data-warehouse EHR for forecasting SARS-CoV-2 hospitalizations in Bordeaux area, France’. In: *JAMIA open* 5.4 (Dec. 2022). DOI: [10.1093/jamiaopen/ooac086](https://doi.org/10.1093/jamiaopen/ooac086). URL: <https://hal.science/hal-03908260> (cit. on p. 6).
- [78] B. P. Hejblum, G. M. Weber, K. P. Liao, N. P. Palmer, S. Churchill, N. Shadick, P. Szolovits, S. N. Murphy, I. Kohane and T. Cai. ‘Probabilistic record linkage of de-identified research datasets with discrepancies using diagnosis codes’. In: *Scientific Data* 6 (2019), p. 180298. DOI: [10.1038/sdata.2018.298](https://doi.org/10.1038/sdata.2018.298). URL: <https://hal.inria.fr/hal-01942279> (cit. on p. 6).
- [79] 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 (cit. on p. 5).
- [80] A. Jarne, D. Commenges, L. Villain, M. Prague, Y. Lévy and R. Thiébaud. ‘Modeling CD4+ T cells dynamics in HIV-infected patients receiving repeated cycles of exogenous Interleukin 7’. In: *The Annals of Applied Statistics* 11.3 (2017), pp. 1593–1616 (cit. on p. 8).
- [81] H. Jin and Z. Liu. ‘A Benchmark for RNA-seq Deconvolution Analysis under Dynamic Testing Environments’. In: *Genome Biology* 22.1 (2021), p. 102. DOI: [10.1186/s13059-021-02290-6](https://doi.org/10.1186/s13059-021-02290-6) (cit. on p. 6).
- [82] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola and L. K. Saul. ‘An introduction to variational methods for graphical models’. In: *Machine learning* 37 (1999), pp. 183–233 (cit. on p. 7).
- [83] E. Kuhn and M. Lavielle. ‘Coupling a stochastic approximation version of EM with an MCMC procedure’. In: *ESAIM: Probability and Statistics* 8 (2004), pp. 115–131 (cit. on p. 7).
- [84] M. Lavielle. *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools*. Chapman and Hall/CRC, 2014. URL: <https://hal.science/hal-01122873> (cit. on p. 5).
- [85] J. Leiner, B. Duan, L. Wasserman and A. Ramdas. ‘Data fission: splitting a single data point’. In: *Journal of the American Statistical Association* (2023), pp. 1–12. DOI: [10.1080/01621459.2023.2270748](https://doi.org/10.1080/01621459.2023.2270748) (cit. on p. 13).
- [86] C. J. Martensen, N. Korsbo, V. Ivaturi and S. Sager. ‘Data-Driven Discovery of Feedback Mechanisms in Acute Myeloid Leukaemia: Alternatives to classical models using Deep Nonlinear Mixed Effect modeling and Symbolic Regression’. In: *bioRxiv* (2024), pp. 2024–06 (cit. on p. 7).
- [87] A. Neufeld, A. Dharamshi, L. L. Gao and D. Witten. ‘Data Thinning for Convolution-Closed Distributions’. In: *Journal of Machine Learning Research* 25.57 (2024), pp. 1–35. URL: <http://jmlr.org/papers/v25/23-0446.html> (cit. on p. 13).
- [88] K. Nganou-Makamdop, A. Talla, A. A. Sharma, S. Darko, A. Ransier, F. Laboune, J. G. Chipman, G. J. Beilman, T. Hoskuldsson, S. Fourati, T. E. Schmidt, S. Arumugam, N. S. Lima, D. Moon, S. Callisto, J. Schoepfoerster, J. Tomalka, P. Mugenyi, F. Ssali, P. Muloma, P. Ssengendo, A. R. Leda, R. K. Cheu, J. K. Flynn, A. Morou, E. Brunet-Ratnasingham, B. Rodriguez, M. M. Lederman, D. E. Kaufmann, N. R. Klatt, C. Kityo, J. M. Brenchley, T. W. Schacker, R. P. Sekaly and D. C. Douek. ‘Translocated Microbiome Composition Determines Immunological Outcome in Treated HIV Infection’. In: *Cell* 184.15 (2021), 3899–3914.e16. DOI: [10.1016/j.cell.2021.05.023](https://doi.org/10.1016/j.cell.2021.05.023) (cit. on p. 6).
- [89] J. T. Ormerod and M. P. Wand. ‘Gaussian variational approximate inference for generalized linear mixed models’. In: *Journal of Computational and Graphical Statistics* 21.1 (2012), pp. 2–17 (cit. on p. 7).
- [90] C. Pasin, I. Balelli, T. Van Effelterre, V. Bockstal, L. Solfrosi, M. Prague, M. Douguilh and R. Thiébaud. ‘Dynamics of the humoral immune response to a prime-boost Ebola vaccine: quantification and sources of variation’. In: *Journal of virology* 93.18 (2019), pp. 10–1128 (cit. on p. 8).
- [91] A. S. Perelson and R. M. Ribeiro. ‘Introduction to modeling viral infections and immunity’. In: *Immunological Reviews* 285.1 (2018), pp. 5–8. DOI: <https://doi.org/10.1111/imr.12700> (cit. on p. 5).
- [92] G. A. Poland, I. G. Ovsyannikova and R. B. Kennedy. ‘Personalized vaccinology: a review’. In: *Vaccine* 36.36 (2018), pp. 5350–5357 (cit. on p. 5).

- [93] M. Prague, D. Commenges, J. Guedj, J. Drylewicz and R. Thiébaud. ‘NIMROD: A program for inference via a normal approximation of the posterior in models with random effects based on ordinary differential equations’. In: *Computer Methods and Programs in Biomedicine* 111.2 (2013), pp. 447–458 (cit. on p. 7).
- [94] M. Prague, D. Commenges, J. M. Gran, B. Ledergerber, J. Young, H. Furrer and R. Thiébaud. ‘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* 73.1 (Mar. 2017), pp. 294–304. DOI: [10.1111/biom.12564](https://doi.org/10.1111/biom.12564). URL: <https://hal.inria.fr/hal-01406614> (cit. on p. 7).
- [95] B. Pulendran. ‘Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology’. In: *Nature Reviews Immunology* 9.10 (2009), pp. 741–7 (cit. on p. 5).
- [96] Z. Qian, W. Zame, L. Fleuren, P. Elbers and M. van der Schaar. ‘Integrating expert ODEs into neural ODEs: pharmacology and disease progression’. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 11364–11383 (cit. on p. 7).
- [97] L. SAS. ‘Monolix version 2021R1’. In: *Antony, France* (2022), <http://lixoft.com/products/monolix/> (cit. on p. 7).
- [98] C. Schubert. ‘Systems immunology: complexity captured’. In: *Nature* 473.7345 (2011), pp. 113–4 (cit. on p. 5).
- [99] J. A. Sinnott, F. Cai, S. Yu, B. P. Hejblum, C. Hong, I. Kohane and 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://doi.org/10.1093/jamia/ocy056). URL: <https://hal.inria.fr/hal-01887930> (cit. on p. 6).
- [100] P. Soret. ‘Régression pénalisée de type Lasso pour l’analyse de données biologiques de grande dimension : application à la charge virale du VIH censurée par une limite de quantification et aux données compositionnelles du microbiote’. Theses. Université de bordeaux, Nov. 2019. URL: <https://hal.inria.fr/tel-02425113> (cit. on p. 6).
- [101] D. Stein, D. O’Connor, C. Blohmke, M. Sadarangani and A. Pollard. ‘Gene Expression Profiles Are Different in Venous and Capillary Blood: Implications for Vaccine Studies’. In: *Vaccine* 34.44 (2016), pp. 5306–5313. DOI: [10.1016/j.vaccine.2016.09.007](https://doi.org/10.1016/j.vaccine.2016.09.007) (cit. on p. 6).
- [102] R. Thiébaud, J. Drylewicz, M. Prague, C. Lacabaratz, S. Beq, A. Jarne, T. Croughs, R.-P. Sekaly, M. M. Lederman, I. Sereti et al. ‘Quantifying and predicting the effect of exogenous interleukin-7 on CD4+ T cells in HIV-1 infection’. In: *PLoS computational biology* 10.5 (2014), e1003630 (cit. on p. 8).
- [103] R. Thiébaud, H. Jacqmin-Gadda, A. Babiker and 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* 24.1 (2005), pp. 65–82 (cit. on p. 7).
- [104] N. Trouvain, L. Pedrelli, T. T. Dinh and X. Hinaut. ‘ReservoirPy: An Efficient and User-Friendly Library to Design Echo State Networks’. In: *Artificial Neural Networks and Machine Learning – ICANN 2020*. Ed. by I. Farkaš, P. Masulli and S. Wermter. Vol. 12397. Cham: Springer International Publishing, 2020, pp. 494–505. DOI: [10.1007/978-3-030-61616-8_40](https://doi.org/10.1007/978-3-030-61616-8_40) (cit. on p. 6).
- [105] L. Wang, J. Cao, J. O. Ramsay, D. Burger, C. Laporte and J. K. Rockstroh. ‘Estimating mixed-effects differential equation models’. In: *Statistics and Computing* 24.1 (2014), pp. 111–121 (cit. on p. 7).
- [106] H. Wu. ‘Statistical methods for HIV dynamic studies in AIDS clinical trials’. In: *Statistical Methods in Medical Research* 14.2 (2005), pp. 171–192 (cit. on p. 7).
- [107] H. Zhang, B. P. Hejblum, G. Weber, N. Palmer, S. Churchill, P. Szolovits, S. Murphy, K. Liao, I. Kohane and T. Cai. ‘ATLAS: an automated association test using probabilistically linked health records with application to genetic studies’. In: *Journal of the American Medical Informatics Association* 28.12 (Dec. 2021), pp. 2582–2592. DOI: [10.1093/jamia/ocab187](https://doi.org/10.1093/jamia/ocab187). URL: <https://hal.inria.fr/hal-03478039> (cit. on p. 6).
- [108] B. J. Zou, M. E. Levine, D. P. Zaharieva, R. Johari and E. B. Fox. ‘Hybrid Square Neural ODE Causal Modeling’. In: *arXiv preprint arXiv:2402.17233* (2024) (cit. on p. 7).