



IN PARTNERSHIP WITH:  
**INSERM**

**Université de Bordeaux**

Activity Report 2014

**Team SISTM**

Statistics in System biology and Translational  
Medicine

RESEARCH CENTER  
**Bordeaux - Sud-Ouest**

THEME  
**Computational Neuroscience and  
Medicine**



## Table of contents

<b>1. Members</b>	<b>1</b>
<b>2. Overall Objectives</b>	<b>1</b>
<b>3. Research Program</b>	<b>2</b>
3.1. Mecanistic modelling	2
3.2. High dimensional data	3
<b>4. Application Domains</b>	<b>3</b>
4.1. Systems Biology and Translational medicine	3
4.2. The case of HIV immunology	4
<b>5. New Software and Platforms</b>	<b>4</b>
5.1. New Software	4
5.2. Upgraded Software	4
5.2.1. NIMROD	4
5.2.2. marqLevAlg	4
5.2.3. VSURF	4
5.2.4. R2GUESS	5
<b>6. New Results</b>	<b>5</b>
6.1. Highlights of the Year	5
6.2. Analysis of purely random forests bias	5
<b>7. Bilateral Contracts and Grants with Industry</b>	<b>5</b>
<b>8. Partnerships and Cooperations</b>	<b>6</b>
8.1. Regional Initiatives	6
8.2. National Initiatives	6
8.2.1. Labex Vaccine Research Institute (VRI)	6
8.2.2. Expert Appraisals	6
8.3. International Initiatives	6
8.4. International Research Visitors	6
8.4.1. Visits of International Scientists	6
8.4.2. Visits to International Teams	6
8.4.2.1. Sabbatical programme	6
8.4.2.2. Research stays abroad	6
<b>9. Dissemination</b>	<b>7</b>
9.1. Promoting Scientific Activities	7
9.1.1. Scientific events organisation	7
9.1.2. Scientific events selection	7
9.1.3. Journal	7
9.1.3.1. Member of the editorial board	7
9.1.3.2. Reviewer	7
9.2. Teaching - Supervision - Juries	8
9.2.1. Teaching	8
9.2.2. Supervision	8
9.2.3. Juries	8
<b>10. Bibliography</b>	<b>8</b>



## Team SISTM

**Keywords:** Statistics, Systems Biology, Stochastic Modeling, Statistical Learning, Epidemiology, Big Data

*The SISTM team is localized in the Carreire area of the Bordeaux University.*

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

## 1. Members

### Research Scientists

Daniel Commenges [INSERM, DR (Emeritus from Sep 2014), HdR]

Laura Richert [Univ. Bordeaux, Assistant Professor (AHU)]

### Faculty Members

Rodolphe Thiebaut [Team leader, Univ. Bordeaux, Prof (PUPH), HdR]

Marta Avalos Fernandez [Univ. Bordeaux, Associate Professor, ADT]

Robin Genuer [Univ. Bordeaux, Associate Professor]

Benoit Liquet [Univ. Bordeaux, Associate Professor, HdR]

### Engineer

Henri Bonnabau [INSERM, Engineer]

### PhD Students

Ana Jarne Munoz [INSERM, from Nov 2012]

Boris Hejblum [Univ. Bordeaux, from Oct 2011]

Marie-Quitterie Picat [Univ. Bordeaux, from Nov 2012]

Perrine Soret [INSERM, from Oct 2014]

### Administrative Assistant

Catherine Cattaert Megrat [Inria]

### Others

Damien Chimits [INSERM, Master Internship, from Apr 2014 to Sep 2014]

Edouard Lhomme [INSERM, Master Internship, from Apr 2014 to Sep 2014]

Melanie Nee [Inria, Master Internship, from Mar 2014 to Sep 2014]

Chloe Pasin [Min. de l'Education Nationale, Master Internship, from Feb 2014]

## 2. Overall Objectives

### 2.1. Overall Objectives

The overall objective of SISTM is to develop statistical methods for the integrative analysis of health data, especially those related to clinical immunology to answer specific questions risen in the application field.

To reach this objective we are developing statistical methods belonging to two main research areas:

Statistical and mechanistic modeling, especially based on ordinary differential equation systems, fitted to population and sparse data

Statistical learning methods in the context of high-dimensional data

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

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

- To understand how immune response is generated with immune interventions (vaccines or interleukines)

- To predict what would be the immune response to a given immune intervention for designing next studies and adapting interventions to individual patients

## 3. Research Program

### 3.1. Mechanistic modelling

When studying the dynamics of a given marker, say the HIV concentration in the blood (HIV viral load), one can for instance use descriptive models summarizing the dynamics over time in term of slopes of the trajectories [48]. These slopes can be compared between treatment groups or according to patients' characteristics. Another way for analyzing these data is to define a mathematical model based on the biological knowledge of what drives HIV dynamics. In this case, it is mainly the availability of target cells (the CD4+ T lymphocytes), the production and death rates of infected cells and the clearance of the viral particles that impact the dynamics. Then, a mathematical model most often based on ordinary differential equations (ODE) can be written [41]. Estimating the parameters of this model to fit observed HIV viral load gave a crucial insight in HIV pathogenesis as it revealed the very short half-life of the virions and infected cells and therefore a very high turnover of the virus, making mutations a very frequent event [40].

Having a good mechanistic model in a biomedical context such as HIV infection opens doors to various applications beyond a good understanding of the data. Global and individual predictions can be excellent because of the external validity of a model based on main biological mechanisms. Control theory may serve for defining optimal interventions or optimal designs to evaluate new interventions [33]. Finally, these models can capture explicitly the complex relationship between several processes that change over time and may therefore challenge other proposed approaches such as marginal structural models to deal with causal associations in epidemiology [32].

Therefore, we postulate that this type of model could be very useful in the context of our research that is in complex biological systems. The definition of the model needs to identify the parameter values that fit the data. In clinical research this is challenging because data are sparse, and often unbalanced, coming from populations of subjects. A substantial inter-individual variability is always present and needs to be accounted as this is the main source of information. Although many approaches have been developed to estimate the parameters of non-linear mixed models [45], [51], [36], [42], [37], [50], the difficulty associated with the complexity of ODE models and the sparsity of the data leading to identifiability issues need further research.

## 3.2. High dimensional data

With the availability of omics data such as genomics (DNA), transcriptomics (RNA) or proteomics (proteins), but also other types of data, such as those arising from the combination of large observational databases (e.g. in pharmacoepidemiology or environmental epidemiology), high-dimensional data have become increasingly common. Use of molecular biological techniques such as Polymerase Chain Reaction (PCR) allows for amplification of DNA or RNA sequences. Nowadays, microarray and Next Generation Sequencing (NGS) techniques give the possibility to explore very large portions of the genome. Furthermore, other assays have also evolved, and traditional measures such as cytometry or imaging have become new sources of big data. Therefore, in the context of HIV research, the dimension of the datasets has much grown in term of number of variables per individual than in term of number of included patients although this latter is also growing thanks to the multi-cohort collaborations such as CASCADE or COHERE organized in the EuroCoord network <sup>1</sup>. As an example, in a recent phase 1/2 clinical trial evaluating the safety and the immunological response to a dendritic cell-based HIV vaccine, 19 infected patients were included. Bringing together data on cell count, cytokine production, gene expression and viral genome change led to a 20 Go database [19]. This is far from big databases faced in other areas but constitutes a revolution in clinical research where clinical trials of hundred of patients sized few hundred of Ko at most. Therefore, more than the storage and calculation capacities, the challenge is the comprehensive analysis of these datasets.

The objective is either to select the relevant information or to summarize it for understanding or prediction purposes. When dealing with high dimensional data, the methodological challenge arises from the fact that datasets typically contain many variables, much more than observations. Hence, multiple testing is an obvious issue that needs to be taken into account [46]. Furthermore, conventional methods, such as linear models, are inefficient and most of the time even inapplicable. Specific methods have been developed, often derived from the machine learning field, such as regularization methods [49]. The integrative analysis of large datasets is challenging. For instance, one may want to look at the correlation between two large scale matrices composed by the transcriptome in the one hand and the proteome on the other hand [38]. The comprehensive analysis of these large datasets concerning several levels from molecular pathways to clinical response of a population of patients needs specific approaches and a very close collaboration with the providers of data that is the immunologists, the virologists, the clinicians...

## 4. Application Domains

### 4.1. Systems Biology and Translational medicine

Biological and clinical researches have dramatically changed because of the technological advances, leading to the possibility of measuring much more biological quantities than previously. Clinical research studies can include now traditional measurements such as clinical status, but also thousands of cell populations, peptides, gene expressions for a given patient. This has facilitated the transfer of knowledge from basic to clinical science (from "bench side to bedside") and vice versa, a process often called "Translational medicine". However, the analysis of these large amounts of data needs specific methods, especially when one wants to have a global understanding of the information inherent to complex systems through an "integrative analysis". These systems like the immune system are complex because of many interactions within and between many levels (inside cells, between cells, in different tissues, in various species). This has led to a new field called "Systems biology" rapidly adapted to specific topics such as "Systems Immunology" [47], "Systems vaccinology" [44], "Systems medicine" [35]. From the statistician point of view, two main challenges appear: i) to deal with the massive amount of data ii) to find relevant models capturing observed behaviors.

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

## 4.2. The case of HIV immunology

The management of HIV infected patients and the control of the epidemics have been revolutionized by the availability of highly active antiretroviral therapies. Patients treated by these combinations of antiretrovirals have most often undetectable viral loads with an immune reconstitution leading to a survival which is nearly the same to uninfected individuals [39]. Hence, it has been demonstrated that early start of antiretroviral treatments may be good for individual patients as well as for the control of the HIV epidemics (by reducing the transmission from infected people) [34]. However, the implementation of such strategy is difficult especially in developing countries. Some HIV infected individuals do not tolerate antiretroviral regimen or did not reconstitute their immune system. Therefore, vaccine and other immune interventions are required. Many vaccine candidates as well as other immune interventions (IL7, IL15) are currently evaluated. The challenges here are multiple because the effects of these interventions on the immune system are not fully understood, there are no good surrogate markers although the number of measured markers has exponentially increased. Hence, HIV clinical epidemiology has also entered in the era of Big Data because of the very deep evaluation at individual level leading to a huge amount of complex data, repeated over time, even in clinical trials that includes a small number of subjects.

# 5. New Software and Platforms

## 5.1. New Software

### 5.1.1. *TcGSA*

An *R* package for the gene set analysis of longitudinal gene expression data sets. Under development, and soon to be available on the CRAN website, this package implements a Time-course Gene Set Analysis method and provides useful plotting functions facilitating the interpretation of the results.

## 5.2. Upgraded Software

### 5.2.1. *NIMROD*

We have written a specific program called NIMROD for estimating parameter of ODE based population models. It has been regularly updated. For instance, we have adapted the program for parallel computing, in collaboration with the MCIA (Mésocentre de calcul intensif Aquitain) facility, which makes available a large computer with more than 3000 cores. This program is described in [43]. Although the program is available on the ISPED website <sup>2</sup>, it is not user-friendly and needs further improvement to be more widely used. By now, the users are the current or previous (Jérémy Guedj, Julia Drylewicz, Mélanie Prague) members of the team and close collaborators (Andrew Yates). Furthermore, as a validation step, it would need a head-to-head comparison with other available softwares. We bet that our program can be very competitive for parameter identification in ODE models with more than two compartments.

### 5.2.2. *marqLevAlg*

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

### 5.2.3. *VSURF*

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

<sup>2</sup><http://etudes.isped.u-bordeaux2.fr/BIOSTATISTIQUE/NIMROD/documentation/html/index.html>.



### 5.2.4. R2GUESS

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

## 6. New Results

### 6.1. Highlights of the Year

A work (described below), in collaboration with M. Davis and R. Tibshirani from Standford University, has been published in the "Proceedings of the National Academy of Sciences" : [8].

Females have generally more robust immune responses than males for reasons that are not well-understood. Here we used a systems analysis to investigate these differences by analyzing the neutralizing antibody response to a trivalent inactivated seasonal influenza vaccine (TIV) and a large number of immune system components, including serum cytokines and chemokines, blood cell subset frequencies, genome-wide gene expression, and cellular responses to diverse in vitro stimuli, in 53 females and 34 males of different ages. We found elevated antibody responses to TIV and expression of inflammatory cytokines in the serum of females compared with males regardless of age. This inflammatory profile correlated with the levels of phosphorylated STAT3 proteins in monocytes but not with the serological response to the vaccine. In contrast, using a machine learning approach, we identified a cluster of genes involved in lipid biosynthesis and previously shown to be up-regulated by testosterone that correlated with poor virus-neutralizing activity in men. Moreover, men with elevated serum testosterone levels and associated gene signatures exhibited the lowest antibody responses to TIV. These results demonstrate a strong association between androgens and genes involved in lipid metabolism, suggesting that these could be important drivers of the differences in immune responses between males and females.

### 6.2. Analysis of purely random forests bias

In collaboration with S. Arlot, we write a research report on some theoretical results about random forests : [30].

Random forests are a very effective and commonly used statistical method, but their full theoretical analysis is still an open problem. As a first step, simplified models such as purely random forests have been introduced, in order to shed light on the good performance of random forests. In this paper, we study the approximation error (the bias) of some purely random forest models in a regression framework, focusing in particular on the influence of the number of trees in the forest. Under some regularity assumptions on the regression function, we show that the bias of an infinite forest decreases at a faster rate (with respect to the size of each tree) than a single tree. As a consequence, infinite forests attain a strictly better risk rate (with respect to the sample size) than single trees. Furthermore, our results allow to derive a minimum number of trees sufficient to reach the same rate as an infinite forest. As a by-product of our analysis, we also show a link between the bias of purely random forests and the bias of some kernel estimators.

## 7. Bilateral Contracts and Grants with Industry

### 7.1. Bilateral Contracts with Industry

Roche Institute, through the Vaccine Research Institute, funding one engineer over 2 years (2012-2014)

Cytheris (now RevImmune), through the ANRS, for the development of IL-7, as this is the only one company able to produce exogenous IL-7 usable in Humans.

## 8. Partnerships and Cooperations

### 8.1. Regional Initiatives

The team have strong links with Bordeaux CHU ("Centre Hospitalier Universitaire").

### 8.2. National Initiatives

#### 8.2.1. *Labex Vaccine Research Institute (VRI)*

There are strong collaborations with immunologists involved in the Labex Vaccine Research Institute (VRI) as RT is leading the Biostatistics/Bioinformatics division.

#### 8.2.2. *Expert Appraisals*

Coordination with Jean Weissenbach of the presidential plan of 100 M€ for "Systems biology" (RT)  
Deputy director of the Institut de Recherche en Santé Publique IRESP (RT)

### 8.3. International Initiatives

#### 8.3.1. *Participation In other International Programs*

RT is participating to the EUROCOORD network on HIV cohort collaborations as :  
a member of the scientific committee of IWHOD International Workshop on HIV Observational Databases from 2013,  
a project leader on defining references for the CD4 count response to antiretrovirals.

### 8.4. International Research Visitors

#### 8.4.1. *Visits of International Scientists*

Following the RHOMEEO project (ANR-BBSRC Systems biology 2007 call, 2007-2011) steered by RT, a strong collaboration has been established with Pr Robin Callard (UCL Immunology) who is visiting the team in Bordeaux one month each year, Andy Yates (Physicists, Glasgow Univ) and Ben Seddon (NIMR, UCL Immunology).

Also, several other international collaboration have been initiated through the Labex:

Steve Self and Peter Gilbert in Seattle (HVTN HIV vaccine Trial Network),  
Marcus Altfeld (Immunologists, Hambourg & Harvard).

This group in collaboration with other teams in Europe is writing a response to the H2020 call PHC 2 – 2015: Understanding diseases: systems medicine.

#### 8.4.2. *Visits to International Teams*

##### 8.4.2.1. *Sabbatical programme*

BL is on sabbatical in Queensland University, Australia.

##### 8.4.2.2. *Research stays abroad*

Chloé Pasin is visiting Steve Self at HVTN, Seattle.

Boris Hejblum visited François Caron at Oxford University, United-Kingdom.

## 9. Dissemination

### 9.1. Promoting Scientific Activities

#### 9.1.1. Scientific events organisation

##### 9.1.1.1. Member of the organizing committee

BMW (Bordeaux Modeling Workshop), a two days workshop was organized (with 30 participants).  
8th French Clinical Epidemiology Conference EPICLIN

#### 9.1.2. Scientific events selection

##### 9.1.2.1. Member of the conference program committee

RT is a member of the scientific committee of IWHOD International Workshop on HIV Observational Databases from 2013,

#### 9.1.3. Journal

##### 9.1.3.1. Member of the editorial board

Lifetime Data Analysis (DC)  
Stat Surveys (DC)  
Journal de la Société Française de Statistique (DC)

##### 9.1.3.2. Reviewer

The members of the team reviewed numerous papers for the following international journals :

AIDS (RT)  
Biometrical (BL)  
Biometrics (DC)  
Annals of applied Statistics (DC)  
Briefings in Bioinformatics (RG)  
Health Services and Outcome Methodology (DC)  
Information Science (RG)  
International Journal of Epidemiology (DC)  
Journal of Multivariate Analysis (RG)  
Journal of the Royal Statistical Society: Series A (DC)  
Machine Learning (RG)  
Neurocomputing (RG)  
Statistics in Medicine (MA, DC, RT)

## 9.2. Teaching - Supervision - Juries

### 9.2.1. Teaching

Master : MA teaches in the two years of the Master of Public Health at ISPED, Univ. Bordeaux, France. Furthermore, she is head of the first year of the master.

Master : DC, teaches occasionally in the Biostatistics specialty of the second year of the Master of Public Health.

Master : RG, teaches in the two years of the Master of Public Health.

BL teaches at the School of Mathematics and Physics (The University of Queensland, Australia).

Master : RT, teaches in the two years of the Master of Public Health, and he is head of the Epidemiology specialty of the second year of the Master of Public Health.

#### E-learning

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

RG teaches in the first year of the e-learning program of the Master of Public Health.

RT is head of the Epidemiology specialty of the second year of the e-learning program of the Master of Public Health, and teaches in it.

### 9.2.2. Supervision

PhD in progress : Ana Jarne, *Modélisation d'interventions sur le système immunitaire pour le traitement et les vaccins contre le VIH*, from Nov 2012, co-directed by Daniel Commenges & Rodolphe Thiébaud

PhD in progress : Boris Hejblum, *Analyse intégrative de données de grande dimension appliquée à la recherche vaccinale*, Oct 2011, co-directed by Rodolphe Thiébaud & François Caron

PhD in progress : Marie-Quitterie Picat, *Analyse des biomarqueurs dans les troubles immunologiques des maladies du système immunitaire*, from Nov 2012, directed by Rodolphe Thiébaud

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

Master internship : Damien Chimits, *Analyse par groupe de gènes de données longitudinales d'expression génique*, from Mar 2014 to Sep 2014, co-directed by Rodolphe Thiébaud & Boris Hejblum

Master internship : Edouard Lhomme, *Modélisation de la dynamique de la réponse immunitaire précoce au vaccin VIH*, from Mar 2014 to Sep 2014, directed by Rodolphe Thiébaud

Master internship : Mélanie Née, *Consommation médicamenteuse et risque d'accident de la route : exploration par simulation de schémas d'études épidémiologiques applicables à partir des données médico-administratives*, from Mar 2014 to Sep 2014, co-directed by Marta Avalos & Ludivine Orriols

Master internship : Chloé Pasin, *Modelling the immune response to HIV vaccine*, from Feb 2014 to Sep 2014, directed by Rodolphe Thiébaud

### 9.2.3. Juries

Members of the team were involved in 6 PhD juries, 2 professorships and 2 HDR.

## 10. Bibliography

### Major publications by the team in recent years

- [1] M. AVALOS, N. D. ADROHER, E. LAGARDE, F. THIESSARD, Y. GRANDVALET, B. CONTRAND, L. ORRIOLS. *Prescription-Drug-Related Risk in Driving: Comparing Conventional and Lasso Shrinkage Logistic Regressions*, in "Epidemiology", 2012, vol. 23, n° 5, pp. 706–712

- [2] D. FURMAN, P. HEJBLUM, N. SIMON, V. JOJIC, L. DEKKER, R. THIÉBAUT, J. TIBSHIRANI, M. DAVIS. *Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination*, in "Proceedings of the National Academy of Sciences", January 2014, vol. 111, n<sup>o</sup> 2, pp. 869-74, <http://dx.doi.org/10.1073/pnas.1321060111>
- [3] B. LIQUET, K.-A. LE CAO, H. HOCINI, R. THIÉBAUT. *A novel approach for biomarker selection and the integration of repeated measures experiments from two assays*, in "BMC bioinformatics", 2012, vol. 13, n<sup>o</sup> 1, 14 p. , <http://dx.doi.org/10.1186/1471-2105-13-325>
- [4] M. PRAGUE, D. COMMENGES, J. DRYLEWICZ, R. THIÉBAUT. *Treatment monitoring of HIV infected patients based on mechanistic models*, in "Biometrics", 2012, vol. 68, n<sup>o</sup> 3, pp. 902-911
- [5] R. THIÉBAUT, J. DRYLEWICZ, M. PRAGUE, C. LACABARATZ, S. BEQ, A. JARNE, T. CROUGHS, R.-P. SEKALY, M. M. LEDERMAN, I. SERETI. *Quantifying and Predicting the Effect of Exogenous Interleukin-7 on CD4+ T Cells in HIV-1 Infection*, in "PLoS computational biology", 2014, vol. 10, n<sup>o</sup> 5, e1003630

## Publications of the year

### Articles in International Peer-Reviewed Journals

- [6] M. AVALOS, L. ORRIOLS, H. POUYES, Y. GRANDVALET, F. THIESSARD, E. LAGARDE. *Variable selection on large case-crossover data: application to a registry-based study of prescription drugs and road traffic crashes*, in "Pharmacoepidemiology and Drug Safety", 2014, pp. 140-51 [DOI : 10.1002/PDS.3539], <https://hal.archives-ouvertes.fr/hal-01099301>
- [7] D. FURMAN, B. P. HEJBLUM, N. SIMON, V. JOJIC, C. L. DEKKER, R. THIEBAUT, R. J. TIBSHIRANI, M. M. DAVIS. *Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination*, in "Proceedings of the National Academy of Sciences U S A", January 2014, vol. 111, n<sup>o</sup> 2, pp. 869-74 [DOI : 10.1073/PNAS.1321060111], <https://hal.inria.fr/hal-01108208>
- [8] D. FURMAN, B. P. HEJBLUM, N. SIMON, V. JOJIC, L. DEKKER, R. THIÉBAUT, J. TIBSHIRANI, M. DAVIS. *Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination*, in "Proceedings of the National Academy of Sciences U S A", January 2014, vol. 111, n<sup>o</sup> 2, pp. 869-74 [DOI : 10.1073/PNAS.1321060111], <http://www.hal.inserm.fr/inserm-00933693>
- [9] J. GAUDART, L. HUIART, P. J. MILLIGAN, R. THIEBAUT, R. GIORGI. *Reproducibility issues in science, is P value really the only answer?*, in "Proceedings of the National Academy of Sciences U S A", May 2014, vol. 111, n<sup>o</sup> 19, E1934 [DOI : 10.1073/PNAS.1323051111], <https://hal.inria.fr/hal-01108195>
- [10] N. HELLARD, M. AVALOS, F. GUIMARAES, J. F. TOUSSAINT, P. DAVID. *Training-Related Risk of Common Illnesses in Elite Swimmers over a Four-Year Period*, in "Medicine and Science in Sports and Exercise", 2014, 1 p. [DOI : 10.1249/MSS.0000000000000461], <https://hal.archives-ouvertes.fr/hal-01099379>
- [11] Y. LÉVY, R. THIÉBAUT, M. MONTES, C. LACABARATZ, L. SLOAN, B. KING, S. PÉRUSAT, C. HARROD, A. COBB, L. K. ROBERTS, M. SURENAUD, C. BOUCHERIE, S. ZURAWSKI, C. DELAUGERRE, L. RICHERT, G. CHÊNE, J. BANCHEREAU, K. PALUCKA. *Dendritic cell-based therapeutic vaccine elicits poly-functional HIV-specific T-cell immunity associated with control of viral load*, in "European Journal of Immunology", August 2014, vol. 44, n<sup>o</sup> 9, pp. 2802-10, <https://hal.inria.fr/hal-01108024>

- [12] K. MOREAU, A. DESSEIX, C. GERMAIN, N. BARTHE, T. BACHELET, D. MOREL, P. MERVILLE, L. COUZI, R. THIEBAUT, P. CHAUVEAU. *Body composition in 98 patients awaiting kidney transplantation*, in "Nutrition", January 2014, vol. 30, n<sup>o</sup> 2, pp. 186-91, <https://hal.inria.fr/hal-01108163>
- [13] A. D. OLSON, L. MEYER, M. PRINS, R. THIEBAUT, D. GURDASANI, M. GUIGUET, M.-L. CHAIX, P. AMORNKUL, A. BABIKER, M. S. SANDHU, K. PORTER, J. F. OKULICZ. *An Evaluation of HIV Elite Controller Definitions within a Large Seroconverter Cohort Collaboration*, in "PLoS ONE", January 2014, vol. 9, n<sup>o</sup> 1, e86719 [DOI : 10.1371/JOURNAL.PONE.0086719], <https://hal.inria.fr/hal-01108192>
- [14] L. ORRIOLS, M. AVALOS, N. MOORE, P. PHILIP, B. DELORME, B. LAUMON, B. GADEGBEKU, L. R. SALMI, E. LAGARDE. *Long-term chronic diseases and crash responsibility: a record linkage study*, in "Accident Analysis and Prevention", 2014, 1 p. [DOI : 10.1016/J.AAP.2014.05.001], <https://hal.archives-ouvertes.fr/hal-01099306>
- [15] M.-Q. PICAT, R. THIÉBAUT, F. LIFERMANN, X. DELBREL, D. ADOUE, L. WITTKOP, A.-L. FAUCHAIS, P. RISPAL, J.-F. MOREAU, J.-F. VIALARD. *T-cell activation discriminates subclasses of symptomatic primary humoral immunodeficiency diseases in adults*, in "BMC Immunology", 2014, vol. 15, n<sup>o</sup> 1, 13 p. [DOI : 10.1186/1471-2172-15-13], <http://www.hal.inserm.fr/inserm-00981320>
- [16] L. RICHERT, A. DOUSSAU, J.-D. LELIÈVRE, V. ARNOLD, V. RIEUX, A. BOUAKANE, Y. LÉVY, G. CHÊNE, R. THIÉBAUT. *Accelerating clinical development of HIV vaccine strategies: methodological challenges and considerations in constructing an optimised multi-arm phase I/II trial design*, in "Trials", February 2014, vol. 15, n<sup>o</sup> 1, 68 p. [DOI : 10.1186/1745-6215-15-68], <http://www.hal.inserm.fr/inserm-00955157>
- [17] E. ROSS, F. TANSER, P. PEI, M.-L. NEWELL, E. LOSINA, R. THIEBAUT, M. WEINSTEIN, K. FREEDBERG, X. ANGLARET, C. SCOTT, F. DABIS, R. WALENSKY. *The Impact of the 2013 WHO Antiretroviral Therapy Guidelines on the Feasibility of HIV Population Prevention Trials*, in "HIV Clinical Trials", January 2014, vol. 15, n<sup>o</sup> 5, pp. 185-98 [DOI : 10.1310/HCT1505-185], <https://hal.inria.fr/hal-01108183>
- [18] R. THIÉBAUT, J. DRYLEWICZ, M. PRAGUE, C. LACABARATZ, S. BEQ, A. JARNE, T. CROUGHS, R.-P. SEKALY, M. M. LEDERMAN, I. SERETI, D. COMMENGES, Y. LÉVY, B. ASQUITH. *Quantifying and Predicting the Effect of Exogenous Interleukin-7 on CD4+T Cells in HIV-1 Infection*, in "PLoS Computational Biology", May 2014, vol. 10, n<sup>o</sup> 5, e1003630 [DOI : 10.1371/JOURNAL.PCBI.1003630], <https://hal.inria.fr/hal-01108204>
- [19] R. THIÉBAUT, P. HEJBLUM, L. RICHERT. *[The analysis of "Big Data" in clinical research.]*, in "Epidemiology and Public Health / Revue d'Epidémiologie et de Santé Publique", January 2014, epub ahead of print [DOI : 10.1016/J.RESPE.2013.12.021], <http://www.hal.inserm.fr/inserm-00933691>
- [20] R. THIÉBAUT, M. PRAGUE, D. COMMENGES. *Mathematical dynamical models for personalized medicine*, in "médecine/sciences", October 2014, vol. 30 Spec No 2, pp. 23-6, <https://hal.inria.fr/hal-01108013>

### International Conferences with Proceedings

- [21] N. FOURET, M. AVALOS, L. WITTKOP, T. RODOLPHE, D. COMMENGES. *Prise en compte de la censure à gauche dans la modélisation de données de grande dimension*, in "46e Journées de Statistique", Rennes, France, June 2014, <https://hal.archives-ouvertes.fr/hal-01099341>

- [22] R. GENUER, J.-M. POGGI, C. TULEAU-MALOT. *VSURF : un package R pour la sélection de variables à l'aide de forêts aléatoires*, in "46èmes Journées de Statistique", Rennes, France, 2014, <https://hal.inria.fr/hal-01096233>
- [23] P. HELLARD, P.-M. LEPRETRE, M. AVALOS, L. SEIFERT, J. F. TOUSSAINT, C. HAUSSWIRTH, P. SAUNDERS, P. DAVID. *Ventilatory, metabolic and kinematic responses in sprint versus distance swimmers*, in "XIth International Symposium on Biomechanics and Medicine in Swimming", Canberra, France, April 2014, <https://hal.archives-ouvertes.fr/hal-01099378>
- [24] C. SCORDIA, M. AVALOS, P. HELLARD. *Comparing methods for summarizing a training load in prediction models of swimming performance*, in "XIIth International Symposium on Biomechanics and Medicine in Swimming", Canberra, Australia, B. MASON (editor), Australian Institute of Sport, April 2014, <https://hal.archives-ouvertes.fr/hal-01099344>

### National Conferences with Proceedings

- [25] N. MÉLANIE, M. AVALOS, L. ORRIOLS, E. LAGARDE. *Etude de l'association entre consommation médicamenteuse et risque d'accident de la route : exploration par simulation de schémas d'études épidémiologiques applicables à partir des données médico-administratives*, in "Journée Nationale Biopharmacie et Santé 2014", Paris, France, November 2014, <https://hal.archives-ouvertes.fr/hal-01099320>

### Conferences without Proceedings

- [26] R. GENUER, J.-M. POGGI, C. TULEAU-MALOT. *VSURF : un package R pour la sélection de variables à l'aide de forêts aléatoires*, in "3èmes Rencontres R", Montpellier, France, 2014, <https://hal.inria.fr/hal-01096237>
- [27] P. SORET, M. AVALOS. *Données longitudinales en grande dimension : état des lieux des packages R*, in "Troisièmes rencontres R", Montpellier, France, June 2014, <https://hal.archives-ouvertes.fr/hal-01099337>
- [28] P. SORET, M. AVALOS. *Méthodes d'apprentissage statistique pour des données longitudinales : une revue systématique*, in "GdR Statistique et Santé", Toulouse, France, October 2014, <https://hal.archives-ouvertes.fr/hal-01099331>

### Scientific Books (or Scientific Book chapters)

- [29] M. AVALOS, Y. GRANDVALET, H. POUYES, L. ORRIOLS, E. LAGARDE. *High-Dimensional Sparse Matched Case-Control and Case-Crossover Data: A Review of Recent Works, Description of an R Tool and an Illustration of the Use in Epidemiological Studies*, in "Computational Intelligence Methods for Bioinformatics and Biostatistics", E. W. ENRICO FORMENTI (editor), Lecture Notes in Computer Science, Springer, 2014, <https://hal.archives-ouvertes.fr/hal-01099313>

### Research Reports

- [30] S. ARLOT, R. GENUER. *Analysis of purely random forests bias*, July 2014, <https://hal.archives-ouvertes.fr/hal-01023596>

### Other Publications

- [31] P. SORET, M. AVALOS. *État de l'art des méthodes de pénalisation pour données répétées dans le temps*, August 2014, Journées MAS de la SMAI, <https://hal.archives-ouvertes.fr/hal-01099367>

## References in notes

- [32] O. AALEN, K. KJETIL ROYSLAND, J. GRAN, B. LEDERGERBER. *Causality, mediation and time: a dynamic viewpoint*, in "Journal of the Royal Statistical Society: Series A (Statistics in Society)", 2007, vol. 175, n<sup>o</sup> 4, pp. 831–861
- [33] F. CASTIGLIONE, B. PICCOLI. *Cancer immunotherapy, mathematical modeling and optimal control*, in "Biometrical Journal", 2007, vol. 247, n<sup>o</sup> 4, pp. 723–32
- [34] R. M. GRANICH, C. F. GILKS, C. DYE, K. M. DE COCK, B. G. WILLIAMS. *Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model*, in "Lancet", 2009, vol. 373, n<sup>o</sup> 9657, pp. 48–57, 0140 6736 English
- [35] L. HOOD, Q. TIAN. *Systems approaches to biology and disease enable translational systems medicine*, in "Genomics Proteomics Bioinformatics", 2012, vol. 10, n<sup>o</sup> 4, pp. 181–5
- [36] Y. HUANG, D. LIU, H. WU. *Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system*, in "Biometrics", 2006, vol. 62, n<sup>o</sup> 2, pp. 413–423
- [37] E. KUHN, M. LAVIELLE. *Maximum likelihood estimation in nonlinear mixed effects models*, in "Computational Statistics & Data Analysis", 2005, vol. 49, n<sup>o</sup> 4, pp. 1020–1038
- [38] K.-A. LE CAO, P. MARTIN, C. ROBERT-GRANIÉ, P. BESSE. *Sparse canonical methods for biological data integration: application to a cross-platform study*, in "BMC bioinformatics", 2009, vol. 10, 34 p.
- [39] C. LEWDEN, D. SALMON, P. MORLAT, S. BEVILACQUA, E. JOUGLA, F. BONNET, L. HERIPRET, D. COSTAGLIOLA, T. MAY, G. CHÊNE. *Causes of death among human immunodeficiency virus (HIV)-infected adults in the era of potent antiretroviral therapy: emerging role of hepatitis and cancers, persistent role of AIDS*, in "International Journal of Epidemiology", 2005, vol. 34, n<sup>o</sup> 1, pp. 121–130, 0300 5771 English
- [40] A. S. PERELSON, A. U. NEUMANN, M. MARKOWITZ, J. M. LEONARD, D. D. HO. *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time*, in "Science", 1996, vol. 271, n<sup>o</sup> 5255, pp. 1582–6
- [41] A. S. PERELSON. *Modelling viral and immune system dynamics*, in "Nature Reviews Immunology", 2002, vol. 2, n<sup>o</sup> 1, pp. 28–36
- [42] J. PINHEIRO, D. BATES. *Approximations to the log-likelihood function in the nonlinear mixed-effects model*, in "Journal of Computational and Graphical Statistics", 1995, vol. 4, n<sup>o</sup> 1, pp. 12–35
- [43] M. PRAGUE, D. COMMENGES, J. GUEDJ, J. DRYLEWICZ, R. THIÉBAUT. *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", 2013, vol. 111, n<sup>o</sup> 2, pp. 447–458
- [44] B. PULENDRAN. *Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology*, in "Nature Reviews Immunology", 2009, vol. 9, n<sup>o</sup> 10, pp. 741–7



- 
- [45] H. PUTTER, S. HEISTERKAMP, J. LANGE, F. DE WOLF. *A Bayesian approach to parameter estimation in HIV dynamical models*, in "Statistics in Medicine", 2002, vol. 21, n<sup>o</sup> 15, pp. 2199–2214
- [46] A. REINER, D. YEKUTIELI, Y. BENJAMINI. *Identifying differentially expressed genes using false discovery rate controlling procedures*, in "Bioinformatics", 2003, vol. 19, n<sup>o</sup> 3, pp. 368–375
- [47] C. SCHUBERT. *Systems immunology: complexity captured*, in "Nature", 2011, vol. 473, n<sup>o</sup> 7345, pp. 113-4
- [48] R. THIÉBAUT, H. JACQMIN-GADDA, A. BABIKER, D. COMMENGES. *Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection*, in "Statistics in Medicine", 2005, vol. 24, n<sup>o</sup> 1, pp. 65-82
- [49] R. TIBSHIRANI. *Regression shrinkage and selection via the lasso*, in "Journal of the Royal Statistical Society: Series B (Statistical Methodology)", 1996, vol. 58, pp. 267–288
- [50] Y. WANG. *Derivation of various NONMEM estimation methods*, in "Journal of Pharmacokinetics and pharmacodynamics", 2007, vol. 34, n<sup>o</sup> 5, pp. 575–593
- [51] H. WU. *Statistical methods for HIV dynamic studies in AIDS clinical trials*, in "Statistical Methods in Medical Research", 2005, vol. 14, n<sup>o</sup> 2, pp. 171–192