

RESEARCH CENTRE

Inria Paris Centre

IN PARTNERSHIP WITH:

INSERM, Université Paris Cité

2023

ACTIVITY REPORT

Project-Team

HEKA

**Health data- and model- driven
approaches for Knowledge Acquisition**

IN COLLABORATION WITH: CENTRE DE RECHERCHE DES CORDELIERS

DOMAIN

Digital Health, Biology and Earth

THEME

**Computational Neuroscience and
Medicine**

Inria

Contents

Project-Team HEKA	1
1 Team members, visitors, external collaborators	2
2 Overall objectives	4
2.1 Context	4
2.2 General aim	5
3 Research program	5
3.1 Axis 1 - Knowledge extraction from clinical data	5
3.1.1 Methods	5
3.2 Axis 2 - Stochastic and data-driven predictive modelling of health trajectories	6
3.2.1 Methods	6
3.3 Axis 3: Data-driven and designs for next generation clinical trials	6
3.3.1 Methods	7
4 Application domains	7
4.1 Multimodal approaches generalizable for several diseases	7
4.1.1 PEPR Digital Health	7
4.1.2 SurvivalGPU – “Using Graphics Processing Units (GPUs) to scale up survival analysis to nation-wide cohorts”	7
4.1.3 Messidore-Inserm BEEP - “Bayesian methods for Early Enriched Platform trials”	8
4.1.4 ANR AT2TA - “Analogies: from Theory to Tools and Applications”	8
4.1.5 HDH BOAS ADHERENCE - “Phenotyping algorithms for the assessment of therapeutic adherence and its clinical consequences in the management of chronic diseases”	8
4.1.6 iDEMO MEDITWIN - Dassault Systems - “Virtual twin for personalised medicine ”(starting in 2024)	9
4.2 Cancer	9
4.2.1 EraPerMed PeCAN (Personalized medicine in CANcer) (EU funding 2020-2023)	9
4.2.2 SIRIC InSiTu - “Insights into cancer: from inflammation to Tumor”	9
4.2.3 Combo - Sanofi - "Evaluating drug combinations in oncology with Real-World Data and state-of-the-art knowledge"	10
4.2.4 RHU OPERANDI - “Optimisation and imPROved Efficacy of targeted RAdioNuclide therapy in Digestive cancers by Imagomics”	10
4.3 Rare diseases and pediatrics	10
4.3.1 EU INVENTS Horizon project - “Innovative designs, extrapolation, simulation methods and evidence-tools for rare diseases addressing regulatory needs” (starting in 2024)	10
4.3.2 EU MSCA Doctoral Networks Orgestra project (starting 2024)	11
4.3.3 CIL LICO- Ciliopathies: group of disorders associated with genetic mutation leading to rare and severe genetic diseases - Projet RHU - Recherche Hospitalo-Universitaire-20118-22	11
4.3.4 BNDMR- Banque Nationale des Maladies Rares	11
4.3.5 Genomic variability transversal program (2018-2023)	12
4.4 Other diseases	12
4.4.1 Antibiotic resistance – FAIR project EU Horizon (on going)	12
4.4.2 Virtual reality (Ongoing)	12
5 Social and environmental responsibility	12
5.1 Impact of research results	12
6 Highlights of the year	12
6.1 Awards	13

7	New software, platforms, open data	13
7.1	New software	13
7.1.1	medkit	13
7.1.2	Pythae	13
7.1.3	Pyraug	13
8	New results	14
8.1	Axis 1	14
8.2	Axis 2	14
8.3	Axis 3	15
9	Bilateral contracts and grants with industry	16
9.1	Bilateral Grants with Industry	16
10	Partnerships and cooperations	17
10.1	European initiatives	17
10.1.1	Horizon Europe	17
10.1.2	H2020 projects	17
10.1.3	Other european programs/initiatives	17
10.2	National initiatives	17
11	Dissemination	17
11.1	Promoting scientific activities	17
11.1.1	Journal	17
11.1.2	Leadership within the scientific community	18
11.1.3	Scientific expertise	18
11.2	Teaching - Supervision - Juries	18
11.2.1	Internal or external Inria responsibilities	18
11.2.2	Interventions	18
12	Scientific production	19
12.1	Publications of the year	19

Project-Team HEKA

Creation of the Project-Team: 2021 November 01

Keywords

Computer sciences and digital sciences

- A3.3. – Data and knowledge analysis
- A3.4. – Machine learning and statistics
- A6.1. – Methods in mathematical modeling
- A6.2. – Scientific computing, Numerical Analysis & Optimization
- A9.1. – Knowledge
- A9.2. – Machine learning
- A9.4. – Natural language processing
- A9.6. – Decision support

Other research topics and application domains

- B2.2. – Physiology and diseases
- B2.3. – Epidemiology
- B2.6. – Biological and medical imaging

1 Team members, visitors, external collaborators

Research Scientists

- Sarah Zohar [Team leader, INSERM, Senior Researcher, HDR]
- Adrien Coulet [INRIA, Associate Professor Detachement, HDR]
- Jean Feydy [INRIA, Researcher]
- Nicolas Garcelon [Fondation Imagine]
- Agathe Guilloux [INRIA, Professor Detachement, HDR]
- Claire Leconte Rives-Lange [INRIA, Advanced Research Position, from Oct 2023]
- Moreno Ursino [INSERM, Researcher]

Faculty Members

- Stephanie Allasonnière [UNIV PARIS CITE, Professor, HDR]
- François Angoulvant [Détachement Univ de Lausanne, Professor, HDR]
- Anita Burgun [UNIV PARIS CITE - APHP, Professor, HDR]
- David Drummond [UNIV PARIS CITE - APHP, Associate Professor]
- Anne-Sophie Jannot [UNIV PARIS CITE - APHP, HDR]
- Sandrine Katsahian [UNIV PARIS CITE - APHP, Professor, HDR]
- Antoine Neuraz [UNIV PARIS CITE - APHP until 31/12/2023]
- Bastien Rance [UNIV PARIS CITE - APHP, Associate Professor]
- Brigitte Sabatier [APHP, Professor, HDR]

Post-Doctoral Fellows

- Nadim Ballout [INRIA, Post-Doctoral Fellow]
- Sarah Berdot [APHP, Post-Doctoral Fellow]
- Sandrine Boulet [INSERM, Post-Doctoral Fellow]
- Xiaoyi Chen [Institut Imagine, until Aug 2023]
- Jong Ho Jhee [INRIA, Post-Doctoral Fellow, from Aug 2023]
- Germain Perrin [APHP, Post-Doctoral Fellow]
- Pierre Sabatier [APHP, until Aug 2023]
- Rosy Tsopra [UNIV PARIS CITE - APHP]

PhD Students

- Safa Alsaidi [INRIA]
- Jean-Baptiste Baitairian [SANOFI, CIFRE, from Apr 2023]
- Nesrine Bannour [Univ Saclay]
- Linus Bleistein [ENS PARIS]
- Tom Boeken [UNIV PARIS CITE - APHP]
- Clément Chadebec [UNIV PARIS CITE, until Sep 2023]
- Pierre Clavier [LIX]
- Lea Comin [CEA, from Oct 2023]
- Aziliz Cottin [DASSAULT SYSTEMES, CIFRE]
- Charles De Ponthaud [UNIV PARIS - CITE, from Nov 2023]
- Benjamin Duputel [EXYSTAT, CIFRE, until Mar 2023]
- Thibaut Fabacher [UNIV STRASBOURG]
- Fabrice Gambaraza [APHP]
- Romain Jaquet [GHNE]
- Emilien Jemelen [Epiconcept, CIFRE, from Nov 2023]
- Enora Laas-Faron [INSTITUT CURIE]
- Judith Lambert [INSERM]
- Ivan Lerner [INSERM]
- Fabien Maury [INSERM, from Oct 2023]
- Juliette Murriss [PIERRE FABRE, CIFRE]
- Lillian Muyama [INRIA]
- Sophie Quennelle [UNIV PARIS, until Nov 2023]
- Alice Rogier [INSERM, until 30/11/2023, then INRIA from 01/12/2023, until Nov 2023]
- Agathe Senellart [UNIV PARIS - CITE]
- Guillaume Serieys [UNIV PARIS - CITE]
- Stylianos Tzedakis [UNIV PARIS - CITE]
- Alexis Van Straaten [UNIV PARIS - CITE]
- Louis Vincent [IMPLICIT, CIFRE]
- Axel Vuorinen [INSERM, from Dec 2023]

Technical Staff

- Deycy Camila Arias Villamil [INRIA, until 31/12/2023, Engineer]
- Armelle Arnoux [AP/HP]
- Olivier Birot [INRIA, until 31/12/2023, Engineer]
- Kim Tam Huynh [Inria, until 31/08/2023, Engineer]
- Louis Pujol [UNIV PARIS - CITE, Engineer]

Interns and Apprentices

- Killian Bakong Epoune [INSERM, Intern, until Jul 2023]
- Lea Comin [AMU, Intern, until Jul 2023]
- Naomi Daval-Pommier [UNIV PARIS - CITE, Intern, from Feb 2023 until Aug 2023]
- Maïsen Hassani [INRIA, Intern, from Apr 2023 until Sep 2023]
- Pierre Olivo-Marin [INSERM, Intern, from Feb 2023 until Aug 2023]
- Mohamed Raissi [INRIA, Intern, from Jun 2023 until Jul 2023]
- Axel Vuorinen [ENSAI, Intern, from Apr 2023 until Sep 2023]

Administrative Assistant

- Meriem Guemair [INRIA]

2 Overall objectives

2.1 Context

Clinicians routinely have to take decisions upon the diagnosis and treatment of complex patients for which clinical guidelines do not provide clear recommendations or do not exist. This is particularly the case for very heterogeneous diseases (e.g., rare diseases or cancer in which clinical manifestations or response to treatment differ frequently from one individual to another) or when patients are seen for a new emerging disease for which no recommendation has been established yet, as it was the case in March, 2020 for the management of COVID-19 infections. Similar situations, leading to delicate decisions, happened in the past but, unfortunately, this experience is hardly taken into account or rationalized for clinicians. Indeed, data related to these past experiences are captured, but until now, these data are not accessible to clinicians and not transformed into high level evidences. This level of evidence is currently only reached by highly controlled analyses (such as controlled clinical trials), for which patients might differ strongly from those treated in routine care, or might have never been seen before in the case of new conditions. Besides, hospital information systems are used at every step of patient care, collecting continuously longitudinal data, both unstructured and structured, including clinical reports, drug prescriptions, physiology, laboratory results, imaging and omics data. These data may even be enriched with medical wearable devices or large-scale claims data such as those of the French Health Data Hub that informs on the global clinical course of patients. Likewise, progresses on artificial intelligence approaches, such as supervised machine learning, have found applications in healthcare enabling for instance clinical decision support systems, patient prioritization, drug repurposing or monitoring drug safety. However, many particularities of health data (e.g., their sensitivity, noisiness, incompleteness, heterogeneity and small volume when one very specific feature or several time points are required), and particularities of health data analysis (e.g., the risk of confoundedness, the need for explanations and fairness) make it challenging to develop tools that are both reliable and usable within hospital work- and patient- flows. For instance, precision medicine requires stratifying smaller and smaller groups

of patients, which may seem contradictory in regards with the general strategy of deep learning that requires large amounts of data to be efficient. Another challenging task is the development of tools that enable gaining knowledge from data agilely, i.e., to update knowledge gain continuously (without compromising on reliability). In summary, methodological developments are required at each step of the health data chain, including: (1) data access, (2) data transformation e.g., via representation learning, (3) data analysis, predictive modelling and knowledge discovery with data- and model- driven approaches, and (4) agile, fast and reliable access to data, implementation of these approaches through applications such as decision support systems, medical devices, next generation clinical trials for the assessment of medical knowledge.

2.2 General aim

The main objective of HeKA is to develop methodologies, tools and their applications in clinics towards a learning health system, i.e., a health system that leverages clinical data collected to extract agilely and reliably novel medical knowledge that, in turn, continuously improves healthcare. Indeed, the availability of EHRs (Electronic Health Records), cohorts and other linked data such as the national Health Data Hub, offers the opportunity to develop models for stratification and prediction with the potential of improving the precision and the personalization of treatments, and thus the quality of healthcare.

3 Research program

The HeKA project-team is following 3 interdependent axes, that are (1) knowledge extraction from clinical data, (2) stochastic and data-driven predictive modelling of health data, and (3) data-driven and designs for next generation clinical trials. These axes can be view and interpreted from a patient care point of view as (1) from patient data to patient representations, (2) from patient data to prediction and decision, and (3) from models to improve patient-related knowledge, respectively. All these axes participate in the development of a learning health system. Axes 1 and 2 can be related to observational studies (either retrospective or prospective) and Axis 3 is related to interventional studies. As a remind, in observational studies the investigator is not acting upon study participants, but only observing relationships between factors and outcomes, while in interventional studies (i.e., clinical trials) the investigator intercedes as part of the study design.

3.1 Axis 1 - Knowledge extraction from clinical data

The development of clinical decision support and statistical predictive models has been historically made by manually selecting and tuning sets of predictive variables. This is a task-dependent and time-consuming operation that neglects most of the available data. Real-world data, such as EHRs or cohorts, offer an access to many variables, even those not initially thought of as predictive. For instance, EHRs consist of structured data, such as demographics, diagnosis, procedures, biological laboratory results, and medication exposures, which can be associated with unstructured data, such as clinical notes, discharge summaries, pathology and imaging reports. In addition, this core EHR data may be complemented with others, including images, omics data, patient-reported outcomes, or conversation transcriptions. However, the use of EHR data for any precision medicine application represents an initial and significant information extraction challenge because of their heterogeneity, incompleteness, and dynamic nature. The aim of this axis is to develop methods and tools for leveraging patients' data in their wide variety and complexity. This encompasses the extraction and transformation of raw data into engineered, featured and learned representations of good quality that will enable or facilitate the development of further clinical decision support and knowledge discovery approaches, as those presented in Axes 2 and 3.

3.1.1 Methods

Methods developed in Axis 1 can be associated with three type of tasks: (i) deep phenotyping, (ii) patient representation and (iii) reasoning with clinical knowledge. (i) Deep phenotyping consists in defining algorithms that enable to identify patients with a particular and potentially complex profile within large healthcare databases. It encompasses the development of natural language processing tools capable of

extracting complex features and their context out of clinical texts; it also includes the ability to consider simultaneously structured and unstructured data of these databases to identify relevant patients. To this aim our methods rely on expert rules, distant-supervision and deep learning language representations. (ii) Regarding patient representation we focus on two distinct kind of representations. The first one is an explicit representation of patients in the form of knowledge graphs, using Semantic Web standards and tools. The second one is a representation of patients within a latent space, using representation learning methods largely inspired from results obtained by deep learning models to learn language representations. (iii) Tasks concerned with reasoning on clinical knowledge are multiple. It encompasses methods to measure patients' similarity between elaborated patients representations (either in the form of knowledge graph or embeddings), hybrid approaches for analogical reasoning and logical and statistical inference.

3.2 Axis 2 - Stochastic and data-driven predictive modelling of health trajectories

The recent availability of high dimensional health data enables the emergence of data-driven models for description, analyses with further possibility to guide clinical decisions. In this high dimensional setting, machine learning-based prediction tasks have been demonstrated to be efficient, although they may not be the best option in every setting. We are interested in these borders between settings, where deep learning approaches fail, but alternatives succeed and reciprocally. Particularly, we are considering borders found in temporal modelling and small-sample settings. Health data provided by EHRs have several specificities among which: (i) patient care trajectories are high dimensional, and (ii) are censored, i.e., data are observed until a certain timepoint. Current models do not succeed in simultaneously tackling these two previous concerns. For example, patient trajectories include comorbidities pathways where one disease may impact the others. Embedding patient trajectories in prognostic models remains a challenge, especially, when low-sample-size high-dimension setting. Machine learning in such setting should be adapted and compared with efficient statistical learning models.

3.2.1 Methods

During this year, we focused on adapting machine learning based methods on the following challenges: signal detection using spatio-temporal data to be able to build health geographical vigilance system, medical image classification in low sample setting, pharmacovigilance signal detection using large database. We proposed a method to model and monitor population distributions over space and time, in order to build an alert system for spatio-temporal data changes based on a new version of the Expectation-Maximization (EM) algorithm to better estimate the number of clusters and their parameters at the same time. We validate this approach on a real data set of positive diagnosed patients to coronavirus disease 2019. We show that our pipeline correctly models evolving real data and detects epidemic changes. We proposed a new method to perform data augmentation in a reliable way in the High Dimensional Low Sample Size setting using a geometry-based variational autoencoder. We validated this approach on a medical imaging classification task where a small number of 3D brain magnetic resonance images are considered and augmented using the proposed framework. We adapted a very classical pharmacoepidemiological method to high dimensional setting, the Weigthed Cumulative Exposure statistical model, which makes it possible to model the temporal relationship between the prescription of a drug and a side effect, by implementing it using Graphics Processing Unit (GPU) programming. We analysed several real life datasets using such implementation and show that it was possible using this adaptation to apply this method to the National Health Insurance Database (SNDS).

3.3 Axis 3: Data-driven and designs for next generation clinical trials

New model-based fundamental researches could help in screening patients and predicting response prior to clinical trials, thanks to biomarkers or EHRs based predictive modelling of patients' response. Digital Medical Devices, are health technologies falling into the definition of medical devices intending to "Prevention, diagnosis, monitoring, treatment or alleviation of disease" (Regulation EU 2017/745) and are mostly based on machine learning algorithms. Among them, learning software as medical devices technologies (SaMD) now have the potential to be translated to patients' care. As such, they should be evaluated in clinical trials. However, by contrast to drugs for which the chemical formula does not change

over time, the performances of a predictive model are constantly enriched by new observational data. The challenge we face now is to be able at any time to ensure the safety and efficacy of SaMD and other advances, for patient care. The objective of this axis is twofold : how clinical trials can help machine learning? And inversely, how can machine learning and other information sources help clinical trials? Accordingly, the first objective is to propose clinical trial methods and designs adapted to continuous learning and adapting tools (DMDs). The second objective, is to produce innovative clinical trial designs acquiring all possible patient-related knowledge: disease and translational models, EHRs, clinical trial data and synthetic patients.

3.3.1 Methods

This year we focused more on the second objective. We developed (i) dose-finding methods that include pharmacokinetics and pharmacodynamics modelling to better choose the appropriate dose-regimen in oncology ; (II) methods based on logistic regression and random effects meta-analysis to syntetise information from multiple early phase trials with several doses ; (iii) methods which allow for analysing disrupted clinical trial by COVID pandemic ; (iv) deep learning models that allow to analyse multistate EHR ...

4 Application domains

4.1 Multimodal approaches generalizable for several diseases

4.1.1 PEPR Digital Health

The PEPR ("Programmes et équipements prioritaires de recherche") Digital Health aims at gathering national multidisciplinary community active in digital health for the development and exploitation of the concept of digital twin in health (started in September 2023).

HeKA's involvement in this PEPR are the following; (i) within project **ShareFAIR**, to learn protocols from clinical data collected along healthcare activity in Electronic Health Records (EHRs) to explicitize the medical decision processes, (steps to reach a particular diagnosis or therapeutic choice) and the management of particular conditions (steps in the management of a particular condition). Protocols extracted from EHRs provide a view on the real-word clinical practice and may then be compared together or with CPG (clinical practice guidelines) which can be seen as more theoretical protocols in that they provide recommendations, or clinical pathways (CP) to standardize clinical practice. It will be applied within **NEUROVASC** in the impact reduction of intracranial aneurysm and stroke, in which we will extract and the proposed clinical pathways, (ii) within **REWIND**, to develop of new mathematical and statistical approaches for the analysis of multimodal multiscale longitudinal data to predict patient's response. These models will be designed, implemented as prototypes and then transferred to an easy-used-well-documented platform where people from diverse communities, in particular physicians, will be able to use them on their own data set, (iii) within **DIGPHAT**, to develop Bayesian modelling of meta-models pathways for the development of digital pharmacological twin; it consists in the analysis of data from omics experiments and selection of relevant covariables and to combine meta-models in pathways to select the most reliable twin model, (iv) within project **M4DI**, to develop a generic method for identifying subgroups of patients with the same phenotype from health databases, using jointly variable correlations and expert data, and to implement it within a computer package, (v) all these previous projects will purpose models or Clinical Decision Support Systems (CDSS) to be translated to clinical practice, however, proof based on data only is not sufficient and it should be evaluated in real life through prospective and interventional clinical trials or studies, this will be done within **SMATCH**. In this project we will propose new methodological paradigms for the clinical evaluation of Digital Medical Devices (DMD) including CDSS and AI based models and algorithms.

4.1.2 SurvivalGPU – "Using Graphics Processing Units (GPUs) to scale up survival analysis to nationwide cohorts"

The recent availability of health insurance databases such as the SNDS opens the door to the detection of adverse drug reactions in the general population. The aim is to generalize the survival analyses usually

carried out during clinical trials on cohorts of $N=1k$ to $10k$ patients to the full French population. This line of research is appealing but poses real methodological challenges. Notably, it requires the development of statistical analysis models that meet the robustness and interpretability requirements of public health physicians while taking full advantage of recent hardware accelerators to scale up to millions of patients per cohort. In this context, our team has been working since 2022 on an efficient re-implementation, on Graphics Processing Units (GPUs), of the standard software tool in the field: the R package "survival". The new "survivalGPU" library (<https://github.com/jeanfeydy/survivalGPU>) leverages recent software tools (PyTorch, PyG, KeOps, reticulate) to bridge the gap between high-performance computing and traditional survival analysis. It now provides a complete re-implementation of the Cox proportional hazards model that is around 100 times faster on GPU than the survival package on CPU. Going further, it supports time-varying drug exposures via the Weighted Cumulative Exposures model and is accessible via an R interface which is fully retro-compatible with that of the survival package. We now intend to perform extensive validation and comparison with other models, prior to pharmaco-epidemiological studies on the SNDS data via the Health Data Hub platform.

4.1.3 Messidore-Inserm BEEP - "Bayesian methods for Early Enriched Platform trials"

The recent pandemic has shown the need of speeding up the clinical trial development of novel or repurposed therapies. Indeed, following the usual drug development paradigm, where clinical trial phases are performed sequentially and separately, the time required to the full process easily exceeds a decade. Our objective is to propose innovative Bayesian enriched "platform" designs for early phase trials, which are adapted to the clinical context and go towards precision medicine. Since we are focusing on early phases of clinical trial, in this setting "platform" cannot be linked to classical RCT. Thus, we aim at defining how "platform" trial should be translated into these early phases. As in the original definition, early platform phases will allow for flexibility, such as adding new arm or stopping treatments for futility (and/or safety in our case). The word enriched refers to the use of new information, or at least not usually used in such early trials, such as positron emission tomography (PET) scan, pharmacokinetics/pharmacodynamics (PK/PD) modelling, mathematical modelling of immune responses, and to the enrichment of the enrolled patient based on their biomarkers. The project is built around workpackages (WPs). In WP1, we develop platform trials in phase 0/I, based on PET-scan; microdosing on several (preclinical) animal species and humans will be adaptively compared, added or deleted to better characterize the extrapolation to human. In WP2, we develop phase I/II dose-finding trial using PK/PD or mechanistic PD models. In WP3, enrichment designs for phase I/II, in survival settings, are proposed when selected biomarkers are available, and the design will be extended in case of combination therapies.

4.1.4 ANR AT2TA - "Analogies: from Theory to Tools and Applications"

Analogical reasoning is a kind of reasoning that is based on finding a common relational system between two situations, exemplars, or domains. In computer science, analogical reasoning can be supported by two main axes of artificial intelligence: knowledge representation and reasoning, and machine learning. The AT2TA projet particularly aims at studying the role that machine learning can play in analogical reasoning; and the HeKA team is in charge of exploring the application of their interplay in the healthcare domain. A PhD student, co-supervised with Inria Paris, IHU Imagine and Université de Lorraine, is learning representations of patients, relying on clinical texts, and study how these representations can first compose analogical propositions, and second serve as bricks to a machine learning architecture for analogical reasoning.

4.1.5 HDH BOAS ADHERENCE - "Phenotyping algorithms for the assessment of therapeutic adherence and its clinical consequences in the management of chronic diseases"

Therapeutic adherence is a complex and multifactorial phenomenon that qualifies the degree to which the patient conforms to medical prescriptions. This phenomenon has both detrimental clinical consequences for the patient (in terms of prognosis and quality of life), and negative economic consequences for the society. One difficulty in the study of adherence is that its measurement is mainly indirect and tainted with many bias. Another difficulty is that adherence is not an end, but should be considered through the lens of patient outcomes. The ADHERENCE project, funded under the Health Data Hub BOAS program

(*Bibliothèque Ouverte d'Algorithmes en Santé*) aims at proposing clearly documented measurements for the level of patient adherence and of its potential clinical outcomes in the context of three chronic diseases (cancer, hypertension and transplantation). One challenge here is to enable these measurements in both EHRs from hospitals and claims from the national insurance database (SNDS).

4.1.6 iDEMO Meditwin - Dassault Systems - "Virtual twin for personalised medicine" (stating in 2024)

Meditwin is a collaborative project funded by BPI ("Banque Publique d'Investissement") with Dassault Systems (leader of the projects), Inria, IHU institutes across France and Medtech startups. The aim of the project is to provide a digital platform relying on virtual twins of individuals who faithfully reproduce their state of health and which make it possible to test different therapeutic options. It will promote interdisciplinarity by facilitating interoperability of multimodal medical data. Our team will use AI approaches to propose Clinical Decision Support Systems (CDSS) in cardiovascular diseases and cancer. We will also develop the clinical trial methodology evaluation these CDSS. In particular, HeKA will develop stratification and classification algorithms, synthetic patient's generators, statistical and mathematical models for multi-modal and multidimensional health data and clinical evaluation methods for the resulted CDSS as Digital Medical Devices.

4.2 Cancer

4.2.1 EraPerMed PeCAN (Personalized medicine in CANcer) (EU funding 2020-2023)

Breast cancer is the most common cancer in women worldwide. The most aggressive type is triple negative breast cancer (TNBC), characterized by a very poor prognosis. New targeted and personalized therapies are urgently required. In this context, the European project PeCAN aims at developing and assessing new "AI algorithms" able to predict treatment response and personalize therapy of TNBC patients. They will propose new approaches for personalization of drug therapy based on a deep molecular characterization of individual tumors and patients and the establishment and use of digital medicine approaches to model effects and side effects of all therapy options. In these projects, our team is in charge of developing multi-scale models that integrate more clinical data in the models (as biomarkers, patients' characteristics, chemo and targeted molecule treatments, biology, clinical outcomes, etc.). We are designing a framework for assessing the clinical performance and safety of the AI algorithms itself, by using real-world data collected in clinical data warehouses and biobanks. These projects are associated with Axes 1-3.

4.2.2 SIRIC InsiTu - "Insights into cancer: from inflammation to Tumor"

To turn scientific knowledge into sustainable healthcare, cancer research must identify who is at risk of cancer, when and in whom a new cancer arises, and how best to treat it and gauge response. Aligned with Europe's Beating Cancer Plan, InsiTu takes on the three challenges of cancer prevention, interception, and treatment in digestive, lung, skin cancer, and heme malignancies. Chronic inflammation is a key cancer niche fostering tumor initiation. Leveraging a transformative Tissue-Hub interfacing diagnostics and research, our program 'From inflammation to clonal emergence and cancer' will unite experts in chronic diseases damage to monitor patients with chronic tissue inflammation and cancer predisposition, mirrored by animal modelling, to understand the critical transition from chronic tissue damage to cancer progression, opening opportunities for prevention and interception. Such longitudinal (and sometimes invasive) interactions between patients and healthcare practitioners can be improved by empowering patients, taking psychic, social and ethical dimensions in consideration. Our program 'Imaging cancer and its environment' will take a different approach to this challenge. Through synergetic interactions with mathematicians and physicists, it will provide novel frameworks for multiscale integration of molecular alterations, cellular processes, and tissue complexity. This effort will result in image-based, non-invasive 'virtual biopsies' as proxies of key biological processes underlying tumor heterogeneity and drug resistance. Along with novel biomarkers such as circulating extracellular vesicles, these virtual biopsies will gauge responses to new therapeutic approaches developed in our third program 'From new targets to new trials'. There, experts in leukemias and skin cancers will use cutting-edge in vivo functional screens and multi-omic interrogation of Tissue-Hub samples to identify new targetable vulnerabilities and develop

next-generation cell-based immunotherapies. To fasten the transfer of these innovations into care, new adaptive clinical trial designs will be engineered.

4.2.3 Combo - Sanofi - "Evaluating drug combinations in oncology with Real-World Data and state-of-the-art knowledge"

Combo is a collaborative project with Industry, national health data platforms and cancer institute: Sanofi Pharma, The Health Data Hub, Centre Léon Berard and Inria-Inserm-HeKA. The objective of the project is to identify promising families of drug combinations in oncology using multisource and multi-modal data modelling and prediction, including RWD (cancer patients' care data from CLCC cancer centre), genomic public databases, literature, clinical trials depository and expert's opinion. Once these combinations will be identified mechanistic models will be used to determine dose-regimen and build dose-finding trial designs for the combinations to be evaluated through formal clinical trials. In this project we lead the following WPs (1) AI based analysis of the multimodal RWD and subgroup discovery for the identification of relevant combinations, (2) Bayesian multi-modal analysis accounting RWD modelling as well as expert's opinion and literature and public clinical platforms AI analysis from Sanofi, and (3) proposing candidates and designs for future phase I designs associated with dose, regimen and associated molecule for the combination on selected family of combinations using preclinical PK/PD model.

4.2.4 RHU OPERANDI - "Optimisation and imProved Efficacy of targeted RADioNuclide therapy in Digestive cancers by Imagomics"

Advanced stage hepatocellular carcinoma (HCC) and gastroenteropancreatic neuroendocrine tumours (GEP-NET) are currently treated with targeted radionuclide therapy (TRT), a highly advanced method that consists of either intra-arterial injections of radioactive microspheres (transarterial radioembolisation - TARE) or targeted peptides radioactively labelled and administered systemically (Peptide Receptor Radionuclide Therapy - PRRT). While highly effective, patient stratification and early identification of responders are currently managed insufficiently due to the lack of pertinent imaging biomarkers, either non-invasive or invasive. Furthermore, therapy-induced DNA damage leads to tumour resistance, reducing TRT efficacy. We aim to overcome those current limits through the OPERANDI project via innovative approaches in engineering, novel imaging biomarkers, and new concepts for DNA repair mechanisms, combined with a fundamental understanding of causal links. Our ambitions go beyond the current state-of-art, embracing even new combinations of drugs and α -emitters to enhance dose localization and efficiency. Methodology will try to understand fundamentally whether current patient management using CT/PET/MRI allows to predict response and survival using cutting edge imaging-based artificial intelligence (AI) approaches in combination with data augmentation techniques to reach statistical significance.

4.3 Rare diseases and pediatrics

4.3.1 EU INVENTS Horizon project - "Innovative designs, extrapolation, simulation methods and evidence-tools for rare diseases addressing regulatory needs" (starting in 2024)

The evaluation of new medicines for rare diseases (RD) including rare paediatric RDs is challenging for several reasons, among which are the small patient sample sizes, heterogeneity of patients and diseases and heterogeneity in disease knowledge. Due to these difficulties, access to effective treatments and the number of treatment options are often limited in RDs. INVENTS aims to provide clinical trial trialists, researchers and regulators with a global framework encompassing methods, workflows and evidence assessment tools to be implemented in orphan and paediatric drug development. Our ambition is to significantly improve the evaluation of evidence and regulatory decision-making through the development and validation of: refined longitudinal model-based diseases trajectories and treatment effect, improved extrapolation models, in silico trials (e.g., virtual patient cohorts), optimised model-based clinical trial designs and evidence synthesis methods. These will be evaluated through simulation studies and tested on extensive data from a range of use cases provided by our industrial partners Roche and Novartis and Real World data (RWD) from RD registry. The INVENTS framework will improve consistency and

efficiency of the drug evaluation process for RD by augmenting clinical evidence without compromising its scientific integrity and providing regulators assessment credibility criteria. At the end of this 5 years project, the European industry will be able to exploit novel and improved clinical trial designs, in silico trials and RWD analysis approaches supporting drug development in RD. The European Medicine Agency and European national regulators (including Health Technology Assessment bodies) will be supplied with a general framework allowing better informed decision making. Most importantly, RD patients will benefit from an increased and faster access to efficacious and safe treatments.

4.3.2 EU MSCA Doctoral Networks Orgestra project (starting 2024)

Organoids experimental models are in vitro 3D cell cultures which can be generated from embryonic stem cells, induced pluripotent stem cells or adult stem cells, and can replicate organs functionally and structurally. Their physiological resemblance to target organs and ability to cryopreserve make organoids a powerful tool for biomedical research and advancing understanding of the mechanisms underlying certain disorders, including rare diseases. The ORGESTRA Joint Doctoral Network will propose innovative organoid technologies for two genetic disorders, i.e., cystic fibrosis and cystinosis. In this project we will supervise 2 PhDs project which will propose statistical development for; (1) linking in silico trials to organoids data and innovative trial design. These designs will incorporate biomarkers-based findings, as organoids, i.e., that reduce unnecessary exposure of patients (screening) or allow drugs to be screened more effectively for non-effectiveness before embarking on human trials. This will be done via a joint doctoral degree with University of Utrecht. (2) Estimand framework involving Bayesian principles on organoids data for clinical trial outcomes and models. The estimands framework will be based on expert's elicitation to understand which questions are more relevant in term of clinical efficacy/toxicity, to select the proper outcomes, to identify the possible intercurrent events and to provide a robust statistical model whose parameters will be estimated under a Bayesian setting. This will be done via a joint doctoral degree with Katholieke Universiteit Leuven.

4.3.3 CIL LICO- Ciliopathies: group of disorders associated with genetic mutation leading to rare and severe genetic diseases - Projet RHU - Recherche Hospitalo-Universitaire- 20118-22

Ciliopathies are a large group of rare and severe genetic diseases caused by ciliary dysfunction associated with clinical and genetic heterogeneity, as well as a lack of knowledge on patients' natural history (i.e., the evolution of the disease). The aim of this RHU-3 project, funded through the Investissements d'avenir program, is to develop innovative, diagnostic, prognostic and tailored therapeutic approaches for patients suffering from ciliopathies to prevent them from developing renal failure. Following this aim, we will develop a mechanistic stratification of ciliopathies, in order to regroup suspected and already diagnosed ciliopathies in a treatment-orientated classification. One goal is to engineer a ready-to-use bio-kit for assessing both diagnosis and prognosis of developing renal alteration, in ciliopathy patients. Finally, personalized therapeutic approaches will be proposed for patients using predictive approaches. We will use methods developed in Axis 1 as well as methods proposed in Axis 2 to model patient trajectory. This project is in collaboration with Imagine Institute, AP-HP, the Medetia company and Ecole Polytechnique.

4.3.4 BNDMR- Banque Nationale des Maladies Rares

The French National Registry for Rare Diseases (BNDMR) is a national tool for epidemiology and public health purposes in the field of rare diseases. In line with the objectives defined by the 2nd and 3rd French National Plan for Rare Diseases, the BNDMR team develops a secure national information system which gathers anonymized clinical data of patients affected by rare diseases in its BNDMR data warehouse. As medical head of the BNDMR, AS Jannot has several research projects strongly connected with HeKA team including CDE.AI and Dromos project. CDE.ai aims to create a set of natural language processing algorithms that will allow the semi-automatic completion of the rare disease minimal data set that is currently completed manually for all patients followed up in the rare disease expert centres. In this project, we will use methods developed in Axe 1 (collaboration with N Garcelon). The DROMOS project is a project that uses the National Data Bank for Rare Diseases by linking it to health insurance data. This matching will allow the description of the care of rare disease patients at the national level for

rare diseases, including the characteristic care of the most frequent rare diseases. We will use methods developed in the form of Axis 2 to model these longitudinal data.

4.3.5 Genomic variability transversal program (2018-2023)

The objective of the INSERM Genomic Variability cross-cutting programme is to understand the role played by genes and their variants on the development of pathologies. This programme, which is based on the longitudinal follow-up of cohorts of individuals and their phenotyping, aims to promote the development of new methods for the analysis of longitudinal data. In this program, we develop in the frame of axis 2 longitudinal approaches to define patients trajectories from truncated data such as those in France National Health Database (SNDS).

4.4 Other diseases

4.4.1 Antibiotic resistance – FAIR project EU Horizon (on going)

The aim of the FAIR project is to evaluate Flagelin aerosol therapy for stimulation of immunity as an alternative treatment against pneumonia with multidrug resistant bacteria. In this project, we are developing a full model using pharmacometrics expertise as well as statistical designs for extrapolation purpose and the design of dose-finding study in healthy volunteers. As written above, in this project, S Zoharvco-lead along with C Kloft (Freie Universitaet Berlin) the WP entitled “Development of a translational modelling and simulation platform for flagellin PK/PD”. The aim of the WP is to propose an optimal design for the first-in-man clinical trials, maximizing knowledge gained from in vitro experimentation, expert knowledge and pre-clinical experiments along the way. By incorporating mechanistic approaches earlier in the development process along with a continuous learning modelling under Bayesian inference, we hope to increase the probability of success of the translation process to the clinical setting and thus, optimizing the statistical design and sample size. This project is in relation with axis 3.

4.4.2 Virtual reality (Ongoing)

Several projects led at the HEGP are currently ongoing to evaluate the analgesia provided by the use of Virtual Reality in different care settings (extracorporeal lithotripsy, after colorectal cancer surgery, and fiberoptic bronchoscopy in critical care). In these projects, not so close but still related to axis-2, we will provide methodological approach and use statistical methods to conclude on the clinical questions, by working closely with all Coordinating Investigators (Prof. D Clause, G Manceau and A Rastello)

5 Social and environmental responsibility

5.1 Impact of research results

Our methods and designs are applied in collaboration with medical research team members at HEGP and Necker hospitals (among others). During the last three years, and in consequence of the COVID-19 sanitary pandemic, we developed and deployed tools for knowledge extraction from clinical text to the central clinical data warehouse of the AP-HP (Entrepôt de Données de Santé, AP-HP) . Through others, these tools enable the recognition of named entities, their context (e.g., negation, personal or family history, hypothesis). They have been widely reused for the development of clinical studies, which led to 13 international publication of clinical epidemiology related to COVID-19. It also serve as a basis for our implication in the 4CE international consortium (<https://www.covidclinical.net/>), which led to 15 publications. These tools have been refactored and their use is facilitate within the medkit library.

6 Highlights of the year

- David Drummond- nomination as MCU-PH in Sept 2023.

- Stéphanie Allasonnière et Anne-Sophie Jannot - nomination at "Mission intergouvernemental sur les données de santé"
- New fundings: PEPR Digital Health (2 projects as PI and 3 as WP leaders), EU Horizon cluster Health as PI, EU MSCA as WP leader, idemo MEDITWIN (please see section Application domains for details)

6.1 Awards

Jean Feydy: "prix science ouverte du logiciel libre de recherche 2023" du ministère de l'Enseignement supérieur et de la Recherche, dans la catégorie "espoir - catégorie documentation"

7 New software, platforms, open data

7.1 New software

7.1.1 medkit

Name: a toolkit for a learning health system

Keywords: Learning health system, Biomedical data, Decision support, Python, Information extraction, Natural language processing, Audio signal processing, Machine learning

Functional Description: This library aims at (1) facilitating the manipulation of healthcare data of various modalities (e.g., structured, text, audio data) for the extraction of relevant features and (2) developing supervised models from these various modalities for decision support in healthcare.

Release Contributions: This release 0.3.1 fixes a pip install issue found for release 0.3.0.

URL: <https://github.com/TeamHeka/medkit>

Contact: Adrien Coulet

Participants: Deycy Camila Arias Villamil, Olivier Birot, Kim-Tam Huynh, Antoine Neuraz, Ivan Lerner, Bastien Rance, Adrien Coulet

7.1.2 Pythae

Keywords: Generative Models, Benchmarking, Reproducibility

Functional Description: This library implements some of the most common (Variational) Autoencoder models under a unified implementation. In particular, it provides the possibility to perform benchmark experiments and comparisons by training the models with the same autoencoding neural network architecture. The feature make your own autoencoder allows you to train any of these models with your own data and own Encoder and Decoder neural networks. It integrates experiment monitoring tools such wandb, mlflow or comet-ml and allows model sharing and loading from the HuggingFace Hub in a few lines of code.

URL: https://github.com/clementchadbec/benchmark_VAE

Contact: Clement Chadbec

7.1.3 Pyraug

Keywords: Generative Models, Data augmentation

Functional Description: This library provides a way to perform Data Augmentation using Variational Autoencoders in a reliable way even in challenging contexts such as high dimensional and low sample size data.

URL: <https://github.com/clementchadebec/pyraug>

Contact: Clement Chadebec

8 New results

The team have generated many results in the last year, here are five illustrations for each axis.

8.1 Axis 1

Neuraz, A., Lerner, I., Birot, O., Arias, C., Han, L., Bonzel, C.L., Cai, T., Huynh K.T., Coulet, A. (2023). TAXN: Translate Align Extract Normalize, a multilingual extraction tool for clinical texts. *Studies in health technology and informatics, To appear.*

Several studies have shown that about 80% of the medical information in an electronic health record is only available through unstructured data. Resources such as medical terminologies in languages other than English are limited and restrain the NLP tools. We propose here to leverage English based resources in other languages using a combination of translation, word alignment, entity extraction and term normalization (TAXN). We implement this extraction pipeline in an opensource library called “medkit”. We demonstrate the interest of this approach through a specific use-case: enriching a phenotypic dictionary for post-acute sequelae in COVID-19 (PASC). TAXN proved to be efficient to propose new synonyms of UMLS terms using a corpus of 70 articles in French with 356 terms enriched with at least one validated new synonym. This study was based on freely available deeplearning models.

Participants: Neuraz A ., Lerner I ., Birot O ., Arias C ., Huynh KT ., Coulet A ..

8.2 Axis 2

We adapted and developed innovative methods based both on machine and deep learning for longitudinal health data in high dimensional setting in different contexts including cluster detection using geographic data, image classification, cut-point detection of prognostic factor, drug side effect detection.

Digital Respiratory Healthcare. Edited by Hilary Pinnock, Vitalii Poberezhets and David Drummond. Book | Published in 2023 DOI: 10.1183/2312508X.erm10223 ISBN (electronic): 978-1-84984-173-3

Respiratory care is undergoing a period of major change as it cautiously begins to embrace digital transformation. Catalysed by the need for remote consultation in the pandemic, time-honoured approaches to delivering care are now being challenged by technology-based initiatives. This Monograph deftly guides the reader through the potential benefits and pitfalls of such change, breaking the discussion down into three areas: technological opportunities and regulatory challenges ; social benefits, challenges and implications; exemplars of digital healthcare. Each chapter reviews contemporary literature and considers not ‘if’ but ‘how’ a digital respiratory future can provide optimal care. The result is an authoritative, balanced guide to developing digital respiratory health.

Participants: Drummond D ..

Lambert J, Leutenegger AL, Jannot AS, Baudot A. Tracking clusters of patients over time enables extracting information from medico-administrative databases. *J Biomed Inform.* 2023 Mar;139:104309. doi: 10.1016/j.jbi.2023.104309. Identifying clusters (i.e., subgroups) of patients from the analysis of medico-administrative databases is particularly important to better understand disease heterogeneity. However, these databases contain different types of longitudinal variables which are measured over different follow-up periods, generating truncated data. It is therefore fundamental to develop clustering approaches that can handle this type of data. We propose here cluster-tracking approaches to identify

clusters of patients from truncated longitudinal data contained in medico-administrative databases. We first cluster patients at each age. We then track the identified clusters over ages to construct cluster-trajectories. We compared our novel approaches with three classical longitudinal clustering approaches by calculating the silhouette score. As a use-case, we analyzed antithrombotic drugs used from 2008 to 2018 contained in the Échantillon Généraliste des Bénéficiaires (EGB), a French national cohort. Our cluster-tracking approaches allow us to identify several cluster-trajectories with clinical significance without any imputation of data. The comparison of the silhouette scores obtained with the different approaches highlights the better performances of the cluster-tracking approaches. The cluster-tracking approaches are a novel and efficient alternative to identify patient clusters from medico-administrative databases by taking into account their specificities.

Participants: Lambert J., Jannot AS ..

Korb-Savoldelli V, Tran Y, Perrin G, Touchard J, Pastre J, Borowik A, Schwartz C, Chastel A, Thervet E, Azizi M, Amar L, Kably B, Arnoux A, Sabatier B. Psychometric Properties of a Machine Learning-Based Patient-Reported Outcome Measure on Medication Adherence: Single-Center, Cross-Sectional, Observational Study. J Med Internet Res. 2023 This study aimed to create and validate a PROM on medication adherence interpreted using an ML approach. This cross-sectional, single-center, observational study was carried out at a French teaching hospital between 2021 and 2022. Eligible patients must have had at least 1 long-term treatment, medication adherence evaluation other than a questionnaire, the ability to read or understand French, an age older than 18 years, and provided their nonopposition. Included adults responded to an initial version of the PROM composed of 11 items, each item being presented using a 4-point Likert scale. The initial set of items was obtained using a Delphi consensus process. Patients were classified as poorly, moderately, or highly adherent based on the results of a medication adherence assessment standard used in the daily practice of each outpatient unit. An ML-derived decision tree was built by combining the medication adherence status and PROM responses. Sensitivity, specificity, positive and negative predictive values (NPVs), and global accuracy of the final 5-item PROM were evaluated. We created an initial 11-item PROM with a 4-point Likert scale using the Delphi process. After item reduction, a decision tree derived from 218 patients including data obtained from the final 5-item PROM allowed patient classification into poorly, moderately, or highly adherent based on item responses. We developed a medication adherence tool based on ML with an excellent NPV. This could allow prioritization processes to avoid referring highly adherent patients to time- and resource-consuming interventions. The decision tree can be easily implemented in computerized prescriber order-entry systems and digital tools in smartphones. External validation of this tool in a study including a larger number of patients with diseases associated with low medication adherence is required to confirm its use in analyzing and assessing the complexity of medication adherence.

Participants: Sabatier, P., Perrin G., JArnoux A ..

8.3 Axis 3

Duputel B, Stallard N, Montestruc F, Zohar S, Ursino M. Using dichotomized survival data to construct a prior distribution for a Bayesian seamless Phase II/III clinical trial. Stat Methods Med Res. 2023 May;32(5):963-977. doi: 10.1177/09622802231160554. Master protocol designs allow for simultaneous comparison of multiple treatments or disease subgroups. Master protocols can also be designed as seamless studies, in which two or more clinical phases are considered within the same trial. They can be divided into two categories: operationally seamless, in which the two phases are separated into two independent studies, and inferentially seamless, in which the interim analysis is considered an adaptation of the study. Bayesian designs are scarcely studied. Our aim is to propose and compare Bayesian operationally seamless Phase II/III designs using a binary endpoint for the first stage and a time-to-event endpoint for the second stage. At the end of Phase II, arm selection is based on posterior (futility)

and predictive (selection) probabilities. The results of the first phase are then incorporated into prior distributions of a time-to-event model. Simulation studies showed that Bayesian operationally seamless designs can approach the inferentially seamless counterpart, allowing for an increasing simulated power with respect to the operationally frequentist design.

Participants: Duputel B., Zohar S., Ursino M..

Yap C, Rekowski J, Ursino M, Solovyeva O, Patel D, Dimairo M, Weir CJ, Chan AW, Jaki T, Mander A, Evans TRJ, Peck R, Hayward KS, Calvert M, Rantell KR, Lee S, Kightley A, Hopewell S, Ashby D, Garrett-Mayer E, Isaacs J, Golub R, Kholmanskikh O, Richards DP, Boix O, Matcham J, Seymour L, Ivy SP, Marshall LV, Hommais A, Liu R, Tanaka Y, Berlin J, Espinasse A, de Bono J. Enhancing quality and impact of early phase dose-finding clinical trial protocols: SPIRIT Dose-finding Extension (SPIRIT-DEFINE) guidance. *BMJ*. 2023 Oct 20;383:e076386. doi: 10.1136/bmj-2023-076386. PMID: 37863491. SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2013 provides guidance for clinical trial protocol writing. However, neither the original guidance nor its extensions adequately cover the features of early phase dose-finding trials. The SPIRIT Dose-finding Extension (DEFINE) statement is a new guideline that provides recommendations for essential items that should be provided in the protocols of these trials. It details extensions to the SPIRIT 2013 guidance, incorporating 17 new items and modifying 15 existing items. The purpose of this guideline is to promote transparency, completeness, reproducibility of methods, and interpretation of early phase dose-finding trial protocols. It is envisioned that the resulting improvements in the design and conduct of early phase clinical trials will ultimately reduce research inefficiencies and inconsistencies, driving transformational advances in clinical care.

Participants: Ursino M..

Hua, W., Mei, H., Zohar, S., Giral, M., and Xu, Y. (2022). Personalized dynamic treatment regimes in continuous time: a Bayesian approach for optimizing clinical decisions with timing. *Bayesian Analysis*, 17(3), 849-878. We developed a two-step Bayesian approach to optimize clinical decisions, estimating personalized optimal dynamic treatment regimes (DTRs), taking into account intervention timing. In the first step, we build a generative model for a sequence of medical interventions with a marked temporal point process (MTPP) where the mark is the assigned treatment or dosage. Then this clinical action model is embedded into a Bayesian joint framework where the other components model clinical observations including longitudinal medical measurements and time-to-event data conditional on treatment histories. In the second step, we proposed a policy gradient method to learn the personalized optimal clinical decision that maximizes the patient survival by interacting the MTPP with the model on clinical observations while accounting for uncertainties in clinical observations learned from the posterior inference of the Bayesian joint model in the first step.

9 Bilateral contracts and grants with industry

9.1 Bilateral Grants with Industry

- **Dassault Systèmes - Cifre:** S. Katsahian and A. Guilloux are supervising Aziliz Cottin on the project: Survival dynamic prediction for personalization of cancer patients' follow-up,
- **Exystat - Cifre:** S. Zohar and M. Ursino are co-supervising Benjamin Duputel on the project: Platform Phase II/III seamless clinical trials Bayesian designs),
- **Pierre Fabre - Cifre:** S. Katsahian is supervising Juliette Murriss on the project: prediction of large-scale recurrent events in digestive cancers,

- **Implicit - Cifre**: S. Allasonnière is supervising Louis Vincent on latent modelling of cardiac time series including external information. A special focus is on missing data and heterogeneous populations.
- **Sanofi - Cifre**: S. Katsahian and A. Guilloux are co-supervising Jean-Baptiste Baitairian on the project: Quantitative Bias Assessment for causal inference.
- **Epiconcept - Cifre**: Sandrine Katsahian and Agathe Guilloux are co-supervising Emilien Jemelen on the project: Evaluation of the contribution of the use of artificial intelligence in the French breast cancer screening program.
- **Combo** - Sanofi - "Evaluating drug combinations in oncology with Real-World Data and state-of-the-art knowledge". Adrien coulet, Moreno Ursino and Sarah Zohar are part of this project. Please see section 4 for further details.

10 Partnerships and cooperations

10.1 European initiatives

10.1.1 Horizon Europe

INVENTS - Sarah Zohar is the PI of this EU Horizon project. Please see section 4 for further details.

10.1.2 H2020 projects

FAIR - Sarah Zohar is a WP leader in this H2020 project. Please see section 4 for further details.

10.1.3 Other european programs/initiatives

The European taskforce lead by EIT Health and French Ministry of Solidarity and Health, for the “harmonization of clinical studies criteria and methodologies in Europe for the evaluation of digital medical devices”. In this taskforce, Sarah Zohar co-lead the WP2 on “Evidence in clinical evaluation” with Corinne Collignon (Head of the Digital Mission at HAS, France) and Barbara Höfgen (Head of the Unit DiGA-Fast-Track at Bfarm, Germany)

EU MSCA Doctoral Networks -Project ORGESTRA "Organoid technologies for disease modeling, drug discovery and development for rare diseases". Sarah Zohar is the WP leader of methodological design for using organoids outcomes and transfer it to clinic.

10.2 National initiatives

ReCAP - Moreno Ursino is the referent of the « Early trials » group. The national network RECaP of Research in Clinical Epidemiology and in Public Health has for purpose to mutualize original research projects and to produce innovations in clinical epidemiology and public health.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Journal

Member of the editorial boards Sarah Zohar is Associated Editor of two scientific journals; "Biometrics" and "Statistics in Biopharmaceutical Research"

Bastien Rance is member of the editorial board of the "International Journal of Medical Informatics".

Reviewer - reviewing activities Members of HeKA are regular reviewers for the following journals: ScientificData, Scientific Reports, SemanticWeb Journal, Journal of the SemanticWeb, Artificial Intelligence in Medicine, statistics in medicine, biometrics, statistics in medical research, etc.

11.1.2 Leadership within the scientific community

Antia Burgun serves as the Representative of the French medical informatics community at the IMIA (International Medical Informatics Association). She also serves as a member of the Executive Board of the Imagine Institute.

11.1.3 Scientific expertise

Sarah Zohar is a voting member at the Cnedimts (“Commission nationale d’évaluation des dispositifs médicaux et des technologies de santé”) at HAS (“Haute Autorité de Santé”).

Sarah Zohar was in the review committee of BpiFrance grant evaluation "évaluation DM à base d'IA et de numérique".

Members of the HeKA team are involved in several ethic committees, such as the Comité d’Ethique de la Recherche APHP.Centre, the EDS AP-HP Comité Scientifique et Ethique. Members of HeKA the scientific board of the ANR generic call Axe H.14 : Interfaces : mathématiques, sciences du numérique – biologie, santé and of the FC3R (French Center for the 3R).

Stéphanie Allassonnière and Anita Burgun hold the Helath Chair at PRAIRIE institute.

Stéphanie Allassonnière is the vice-president for valorization of Université Paris Cité

11.2 Teaching - Supervision - Juries

Anne Sophie Jannot and Sandrine Katsahian lead the speciality “Big Data in Health” in the Master of Science of Public Health at the Paris Cité University.

Anne Sophie Jannot co-leads the quantitative biomedicine course, which is part from the medical degree course.

Anne Sophie Jannot leads a professional degree of “health data reuse” at the Paris Cité University in collaboration with Marseille and Bordeaux Universities.

Sandrine Katsahian is responsible for the PRIME department dedicated to the Research, Innovation, Digital Medicine located in the HEGP in APHP.Centre.

Moreno URSINO co-leads the course “Science des données” in the L2SIAS at the Paris Cité University.

Stéphanie Allassonniere within the MVA (Mathematics, Vision, Learning) M2 run a course focusing on the analysis of real-life health data.

Stéphanie Allassonniere coordinates the SMPS Bioentrepreneur (Université Paris Cité)

Jean Feydy is involved in the digital health track of the EPITA school of software engineer

Members of the HeKA team are in charge of two majors of the Master in Public Health of the Université Paris Cité: one on Biomedical Informatics (IBM) and one on Data Science in Healthcare (DMS).

11.2.1 Internal or external Inria responsibilities

Stéphanie Allassonnière was the chair of the CRCN-ISFP INRIA Lyon committee.

11.2.2 Interventions

HeKA members are regularly invited to present their research to the general public.

12 Scientific production

12.1 Publications of the year

International journals

- [1] C. Bodemer, L. Soussand, A. Sandrin, A. Khatim, A. Sauvestre, N. Elarouci and A.-s. Jannot. 'French data on the epidemiology and expert healthcare network for epidermolysis bullosa'. In: *Journal of the European Academy of Dermatology and Venereology* 37.5 (17th Jan. 2023). DOI: [10.1111/jdv.18826](https://doi.org/10.1111/jdv.18826). URL: <https://inria.hal.science/hal-04362512>.
- [2] T. Boeken, J. Feydy, A. Lecler, P. Soyer, A. Feydy, M. Barat and L. Duron. 'Artificial intelligence in diagnostic and interventional radiology: Where are we now?' In: *Diagnostic and Interventional Imaging* 104.1 (Jan. 2023), pp. 1–5. DOI: [10.1016/j.diii.2022.11.004](https://doi.org/10.1016/j.diii.2022.11.004). URL: <https://hal.science/hal-03949288>.
- [3] J. F. van Boven, D. Drummond, A. H. Chan, M. Hew, C. Y. Hui, I. Adejumo, I. Cano, K. Hansen, V. Poberezhets, R. W. Costello and H. Pinnock. 'ERS "CONNECT" Clinical Research Collaboration – moving multiple digital innovations towards connected respiratory care: addressing the overarching challenges of whole systems implementation'. In: *European Respiratory Journal* 62.5 (Nov. 2023), p. 2301680. DOI: [10.1183/13993003.01680-2023](https://doi.org/10.1183/13993003.01680-2023). URL: <https://hal.science/hal-04339221>.
- [4] S. Calderazzo, S. Tarima, C. Reid, N. Flournoy, T. Friede, N. Geller, J. Rosenberger, N. Stallard, M. Ursino, M. Vandemeulebroecke, K. van Lancker and S. Zohar. 'Coping with Information Loss and the Use of Auxiliary Sources of Data: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions'. In: *Statistics in Biopharmaceutical Research* (26th June 2023), pp. 1–17. DOI: [10.1080/19466315.2023.2211023](https://doi.org/10.1080/19466315.2023.2211023). URL: <https://hal.science/hal-04336804>.
- [5] G. Cheminet, A. Brunetti, D. Khimoud, B. Ranque, A. Michon, E. Flamarion, J. Pouchot, A.-s. Jannot and J.-B. Arlet. 'Acute chest syndrome in adult patients with sickle cell disease: The relationship with the time to onset after hospital admission'. In: *British Journal of Haematology* 201.6 (25th Mar. 2023), pp. 1229–1238. DOI: [10.1111/bjh.18777](https://doi.org/10.1111/bjh.18777). URL: <https://inria.hal.science/hal-04362502>.
- [6] M. Cosenza-Contreras, A. Schäfer, J. Sing, L. Cook, M. Stillger, C.-Y. Chen, J. Villacorta Hidalgo, N. Pinter, L. Meyer, T. Werner, D. Bug, Z. Haberl, O. Kübeck, K. Zhao, S. Stei, A. V. Gafencu, R. Ionita, F. Brehar, J. Ferrer-Lozano, G. Ribas, L. Cerdá-Alberich, L. Martí-Bonmatí, C. Nimsky, A. van Straaten, M. Biniossek, M. Föll, N. Cabezas-Wallscheid, J. Büscher, H. Röst, A. Arnoux, J. Bartsch and O. Schilling. 'Proteometabolomics of initial and recurrent glioblastoma highlights an increased immune cell signature with altered lipid metabolism'. In: *Neuro-Oncology* (26th Oct. 2023). DOI: [10.1093/neuonc/noad208](https://doi.org/10.1093/neuonc/noad208). URL: <https://hal.science/hal-04343649>.
- [7] A. Decoux, L. Duron, P. Habert, V. Roblot, E. Arsovic, G. Chassagnon, A. Arnoux and L. Fournier. 'Comparative performances of machine learning algorithms in radiomics and impacting factors'. In: *Scientific Reports* 13.1 (28th Aug. 2023), p. 14069. DOI: [10.1038/s41598-023-39738-7](https://doi.org/10.1038/s41598-023-39738-7). URL: <https://hal.science/hal-04343652>.
- [8] R. Degrave, J. Murriss, A. Charles-Nelson, O. Hermine, R. Porcher, P. Ravaut, X. Mariette, P.-L. Tharaux, M. Resche-Rigon, O. Sanchez, S. Katsahian and J.-B. Arlet. 'Risk factors for thromboembolic events in patients hospitalized for COVID-19 pneumonia in a general ward and requiring treatment with oxygen'. In: *Postgraduate Medical Journal* (17th Nov. 2023). DOI: [10.1093/postmj/qgad104](https://doi.org/10.1093/postmj/qgad104). URL: <https://hal.science/hal-04348667>.
- [9] M.-A. Delbarre, F. Girardon, L. Roquette, P. Blanc-Durand, M.-A. Hubaut, É. Hachulla, F. Semah, D. Huglo, N. Garcelon, E. Marchal, I. El Esper, C. Tribouilloy, N. Lamblin, P. Duhaut, J. Schmidt, E. Itti and T. Damy. 'Deep Learning on Bone Scintigraphy to Detect Abnormal Cardiac Uptake at Risk of Cardiac Amyloidosis'. In: *JACC: Cardiovascular Imaging* (Apr. 2023). DOI: [10.1016/j.jcmg.2023.01.014](https://doi.org/10.1016/j.jcmg.2023.01.014). URL: <https://u-picardie.hal.science/hal-04082960>.

- [10] A. Dhote, S. Tzedakis, C. Hobeika, U. Marchese, R. Coriat, A. Nassar, A. Beaufrère, T. Codjia, C. Goumard, O. Scatton, A. Dohan, M. Ronot, B. Terris, M. Lesurtel, S. Dokmak, F. Cauchy and D. Fuks. 'Impact of induction therapy in locally advanced intrahepatic cholangiocarcinoma'. In: *EJSO - European Journal of Surgical Oncology* (May 2023). DOI: [10.1016/j.ejso.2023.05.014](https://doi.org/10.1016/j.ejso.2023.05.014). URL: <https://hal.science/hal-04172878>.
- [11] D. Drummond and A. Gonsard. 'Télésurveillance dans l'asthme pédiatrique : vers des systèmes cyber-physiques'. In: *Perfectionnement en Pédiatrie* 6.4 (Dec. 2023), pp. 244–249. DOI: [10.1016/j.perped.2023.10.009](https://doi.org/10.1016/j.perped.2023.10.009). URL: <https://hal.science/hal-04339220>.
- [12] D. Drummond, J. Roukema and M. Pijnenburg. 'Home monitoring in asthma: towards digital twins'. In: *Current Opinion in Pulmonary Medicine* 29.4 (27th July 2023), pp. 270–276. DOI: [10.1097/MCP.0000000000000963](https://doi.org/10.1097/MCP.0000000000000963). URL: <https://hal.science/hal-04339229>.
- [13] D. Drummond, P. Whelan and M. Sperrin. 'Association between nebuliser therapies adherence and visit-to-visit variability of FEV1 in patients with cystic fibrosis'. In: *Journal of Cystic Fibrosis* 22.4 (July 2023), pp. 702–705. DOI: [10.1016/j.jcf.2023.03.006](https://doi.org/10.1016/j.jcf.2023.03.006). URL: <https://hal.science/hal-04339233>.
- [14] B. Duputel, N. Stallard, F. Montestruc, S. Zohar and M. Ursino. 'Using dichotomized survival data to construct a prior distribution for a Bayesian seamless Phase II/III clinical trial'. In: *Statistical Methods in Medical Research* 32.5 (May 2023), pp. 963–977. DOI: [10.1177/09622802231160554](https://doi.org/10.1177/09622802231160554). URL: <https://hal.science/hal-04126394>.
- [15] L. Farah, J. Murriss, I. Borget, A. Guilloux, N. Martelli and S. I. Katsahian. 'Assessment of Performance, Interpretability, and Explainability in Artificial Intelligence–Based Health Technologies: What Healthcare Stakeholders Need to Know'. In: *Mayo Clinic Proceedings: Digital Health* 1.2 (June 2023), pp. 120–138. DOI: [10.1016/j.mcpdig.2023.02.004](https://doi.org/10.1016/j.mcpdig.2023.02.004). URL: <https://hal.science/hal-04348677>.
- [16] C. Faviez, M. Talmatkadi, P. Foulquié, A. Mebarki, S. Schück, A. Burgun and X. Chen. 'Assessment of the Early Detection of Anosmia and Ageusia Symptoms in COVID-19 on Twitter: Retrospective Study'. In: *JMIR Infodemiology* 3 (2023), e41863. DOI: [10.2196/41863](https://doi.org/10.2196/41863). URL: <https://inria.hal.science/hal-04227117>.
- [17] L. Gaucher, P. Sabatier, S. Katsahian and A.-S. Jannot. 'Pharmacovigilance studies without a priori hypothesis: systematic review highlights inappropriate multiple testing correction procedures'. In: *Journal of Clinical Epidemiology* 162 (Oct. 2023), pp. 127–134. DOI: [10.1016/j.jclinepi.2023.08.010](https://doi.org/10.1016/j.jclinepi.2023.08.010). URL: <https://hal.science/hal-04348670>.
- [18] A. Gonsard, R. Aboutaam, B. Prévost, C. Roy, A. Hadchouel, N. Nathan, J. Taytard, A. Pirojoc, C. Delacourt, S. Wanin and D. Drummond. 'Children's views on artificial intelligence and digital twins for the daily management of their asthma: a mixed-method study'. In: *European Journal of Pediatrics* 182.2 (Feb. 2023), pp. 877–888. DOI: [10.1007/s00431-022-04754-8](https://doi.org/10.1007/s00431-022-04754-8). URL: <https://hal.science/hal-03993835>.
- [19] E. Gras, A. Jean, V. Rocher, Y. Tran, S. Katsahian, D. Jouclas, C. Dano, J. Cedile, D. Manar, N. Kassis Chikhani, J. Le Guen, C. Patas D'illiers and D. Lebeau. 'Incidence of and risk factors for local complications of peripheral venous catheters in patients older than 70 years: Empirical research quantitative'. In: *Journal of Clinical Nursing* 32.15-16 (Aug. 2023), pp. 5000–5009. DOI: [10.1111/jocn.16732](https://doi.org/10.1111/jocn.16732). URL: <https://hal.science/hal-04348675>.
- [20] P. Habert, A. Decoux, L. Chermati, L. Gibault, P. Thomas, A. Varoquaux, F. Le Pimpec-Barthes, A. Arnoux, L. Juquel, K. Chaumoitre, S. Garcia, J.-Y. Gaubert, L. Duron and L. Fournier. 'Best imaging signs identified by radiomics could outperform the model: application to differentiating lung carcinoid tumors from atypical hamartomas'. In: *Insights into Imaging* 14.1 (19th Sept. 2023), p. 148. DOI: [10.1186/s13244-023-01484-9](https://doi.org/10.1186/s13244-023-01484-9). URL: <https://hal.science/hal-04343651>.
- [21] B. Hansel, P. Sabatier, D. Kadouch, T. Poghosyan, K. Arapis, C. Rives-Lange, S. Czernichow, A.-S. Jannot and C. Carette. 'Bariatric Surgery Should Be Proposed in Certain Septuagenarian Patients with Obesity'. In: *Obesity Surgery* 33.8 (22nd June 2023), pp. 2608–2610. DOI: [10.1007/s11695-023-06655-7](https://doi.org/10.1007/s11695-023-06655-7). URL: <https://inria.hal.science/hal-04362497>.

- [22] E. Kempf, M. Vaterkowski, D. Leprovost, N. Griffon, D. Ouagne, S. Breant, P. Serre, A. Mouchet, B. Rance, G. Chatellier, A. Bellamine, M. Frank, J. Guerin, X. Tannier, A. Livartowski, M. Hilka and C. Daniel. 'How to Improve Cancer Patients Enrollment in Clinical Trials From Real-Life Databases Using the Observational Medical Outcomes Partnership Oncology Extension: Results of the PENELOPE Initiative in Urologic Cancers'. In: *JCO Clinical Cancer Informatics* 7 (May 2023). DOI: [10.1200/CCI.22.00179](https://doi.org/10.1200/CCI.22.00179). URL: <https://hal.science/hal-04107871>.
- [23] V. Korb-Savoldelli, Y. Tran, G. Perrin, J. Touchard, J. Pastre, A. Borowik, C. Schwartz, A. Chastel, E. Thervet, M. Azizi, L. Amar, B. Kably, A. Arnoux and B. Sabatier. 'Psychometric Properties of a Machine Learning-Based Patient-Reported Outcome Measure on Medication Adherence: Single-Center, Cross-Sectional, Observational Study'. In: *Journal of Medical Internet Research* 25 (16th Oct. 2023), e42384. DOI: [10.2196/42384](https://doi.org/10.2196/42384). URL: <https://hal.science/hal-04343657>.
- [24] J. Lambert, A.-L. Leutenegger, A.-S. Jannot and A. Baudot. 'Tracking clusters of patients over time enables extracting information from medico-administrative databases'. In: *Journal of Biomedical Informatics* 139 (Mar. 2023), p. 104309. DOI: [10.1016/j.jbi.2023.104309](https://doi.org/10.1016/j.jbi.2023.104309). URL: <https://hal.science/hal-04027783>.
- [25] A. Lansier, C. Bourillon, C.-A. Cuénod, E. Ragot, A. Follin, S. Hamada, O. Clément, P. Soyer and A.-S. Jannot. 'CT-based diagnostic algorithm to identify bowel and/or mesenteric injury in patients with blunt abdominal trauma'. In: *European Radiology* 33 (2023), pp. 1918–1927. DOI: [10.1007/s00330-022-09200-9](https://doi.org/10.1007/s00330-022-09200-9). URL: <https://hal.science/hal-03883935>.
- [26] V. Maisons, Y. Ramdani, A. Hankard, C. Messiaen, A.-S. Jannot, B. Sautenet, J.-M. Halimi, F. Maillot, É. Pillebout and A. Audemard-Verger. 'New insights into epidemiological data and impact of the COVID-19 pandemic on IgA vasculitis in children and adults: a French nationwide cohort'. In: *Rheumatology International* 43.10 (12th July 2023), pp. 1791–1798. DOI: [10.1007/s00296-023-05387-2](https://doi.org/10.1007/s00296-023-05387-2). URL: <https://inria.hal.science/hal-04362465>.
- [27] R. Marcilly, W.-Y. Zheng, P. Quindroit, S. Pelayo, S. Berdot, B. Charpiat, J. Corny, S. Drouot, P. Frery, G. Leguelinel-Blache, L. Mondet, A. Potier, L. Robert, L. Ferret and M. Baysari. 'Comparison of the validity, perceived usefulness, and usability of I-MeDeSA and TEMAS, two tools to evaluate alert system usability'. In: *International Journal of Medical Informatics* 175 (July 2023), p. 105091. DOI: [10.1016/j.ijmedinf.2023.105091](https://doi.org/10.1016/j.ijmedinf.2023.105091). URL: <https://hal.science/hal-04097607>.
- [28] P. Martin, C. Alberti, A. Bourmaud, A. Burgun and E. de La Rochebrochard. 'Communauté interactive et participative en ligne, menée avec les pairs, comme outil de promotion de la santé sexuelle et reproductive des adolescents et jeunes adultes: la recherche-action Sexpairs'. In: *Epidemiology and Public Health = Revue d'Epidémiologie et de Santé Publique* 71.Supplement 2 (May 2023), pp. 6–7. DOI: [10.1016/j.respe.2023.101609](https://doi.org/10.1016/j.respe.2023.101609). URL: <https://hal.science/hal-04208892>.
- [29] M. Mayenga, N. Falvo, I. Mahé, A.-S. Jannot, B. Gazeau, G. Meyer, N. Gendron, O. Sanchez, S. Djennaoui and B. Planquette. 'Cancer-Associated Thrombosis on Bevacizumab: Risk of Recurrence and Bleeding When Bevacizumab Is Stopped or Continued'. In: *Cancers* 15.15 (31st July 2023), p. 3893. DOI: [10.3390/cancers15153893](https://doi.org/10.3390/cancers15153893). URL: <https://inria.hal.science/hal-04362460>.
- [30] A. Mouazer, N. Boudegzdame, K. Sedki, R. Tsopra and J.-B. Lamy. 'Model and Strategy for Predicting and Discovering Drug-Drug Interactions'. In: *Studies in Health Technology and Informatics. Studies in Health Technology and Informatics* (18th May 2023). DOI: [10.3233/SHTI230248](https://doi.org/10.3233/SHTI230248). URL: <https://hal.science/hal-04162795>.
- [31] J. Murris, A. Charles-Nelson, A. Tadmouri Sellier, A. Lavenu and S. Katsahian. 'Towards filling the gaps around recurrent events in high dimensional framework: a systematic literature review and application'. In: *Biostatistics & Epidemiology* 7.1 (22nd Nov. 2023). DOI: [10.1080/24709360.2023.2283650](https://doi.org/10.1080/24709360.2023.2283650). URL: <https://hal.science/hal-04320229>.
- [32] T. Pichon, C. Messiaen, L. Soussand, C. Angin, A. Sandrin, N. Elarouci and A.-S. Jannot. 'Overview of patients' cohorts in the French National rare disease registry'. In: *Orphanet Journal of Rare Diseases* 18.1 (3rd July 2023), p. 176. DOI: [10.1186/s13023-023-02725-2](https://doi.org/10.1186/s13023-023-02725-2). URL: <https://inria.hal.science/hal-04362496>.

- [33] S. Quennelle and D. Bonnet. 'Pediatric heart failure with preserved ejection fraction, a review'. In: *Frontiers in Pediatrics* 11 (2nd Aug. 2023). DOI: [10.3389/fped.2023.1137853](https://doi.org/10.3389/fped.2023.1137853). URL: <https://hal.science/hal-04191461>.
- [34] C. Rives-Lange, S. Czernichow and A.-S. Jannot. 'Is Maternal Metabolic Bariatric Surgery the Best Solution to Tackle the Childhood Obesity Pandemic?—Reply'. In: *JAMA Surgery* 158.8 (1st Aug. 2023), p. 892. DOI: [10.1001/jamasurg.2023.0931](https://doi.org/10.1001/jamasurg.2023.0931). URL: <https://inria.hal.science/hal-04362499>.
- [35] M. Sapoval, C. Querub, H. Pereira, O. Pellerin, T. Boeken, A. Di Gaeta, M. A. Ahmar, M.-M. Lefevre-Colau, C. Nguyen, C. Daste, M. Lacroix, J.-D. Laredo, B. Sabatier, N. Martelli, G. Chatellier, C. Déan and F. Rannou. 'Genicular artery embolization for knee osteoarthritis: Results of the LipioJoint-1 trial'. In: *Diagnostic and Interventional Imaging* (Dec. 2023). DOI: [10.1016/j.diii.2023.12.003](https://doi.org/10.1016/j.diii.2023.12.003). URL: <https://hal.science/hal-04349672>.
- [36] S. Schmartz, A. Welfringer-Morin, M. Le Bourgeois, C. Delacourt, L. Berteloot, S. H. Rabia, N. Bellon, F. Moulin, M. Pouletty, N. Garcelon, C. Bodemer and D. Drummond. 'Respiratory complications in pediatric epidermal necrolysis: A retrospective study of 22 cases'. In: *Journal of The American Academy of Dermatology* 89.4 (Oct. 2023), pp. 857–859. DOI: [10.1016/j.jaad.2023.05.092](https://doi.org/10.1016/j.jaad.2023.05.092). URL: <https://hal.science/hal-04339228>.
- [37] F. Schortgen, C. Tabra Osorio, D. Carpentier, M. Henry, P. Beuret, G. Lacave, G. Simon, P.-Y. Blanchard, T. Gobe, A. Guillon, L. Bitker, G. Duhommet, J.-P. Quenot, M. Le Meur, S. Jochmans, F. Dubouloz, N. Mainguy, J. Saletes, T. Creutin, P. Nicolas, J. Senay, A.-L. Berthelot, D. Rizk, D. Tran Van, A. Riviere, S. B. Heili-Frades, J. Nunes, N. Robquin, S. Lhotellier, S. Ledochowski, A. Guénégon-Arnoux and A. Constan. 'Fluid Intake in Critically Ill Patients: The “Save Useless Fluids For Intensive Resuscitation” Multicenter Prospective Cohort Study'. In: *Critical Care Medicine* (1st Nov. 2023). DOI: [10.1097/CCM.0000000000006091](https://doi.org/10.1097/CCM.0000000000006091). URL: <https://hal.science/hal-04343611>.
- [38] D. Sciard, K. Xu, A. Soulier, G. Dufour, M. Ursino, C. Alberti and M. Beaussier. 'Postoperative analgesia after surgical repair of distal radius fracture: a randomized comparison between distal peripheral nerve blockade and surgical site infiltration'. In: *Minerva Anestesiologica* 89.10 (Oct. 2023). DOI: [10.23736/s0375-9393.23.16956-2](https://doi.org/10.23736/s0375-9393.23.16956-2). URL: <https://hal.science/hal-04336790>.
- [39] D. Smadja, A.-S. Jannot, A. Philippe, E. Lu, J. Rancic, O. Sanchez, R. Chocron, N. Gendron and J.-L. Diehl. 'Circulating Von Willebrand factor: a consistent biomarker predicting in-hospital mortality across different waves of the COVID-19 pandemic'. In: *Angiogenesis* (9th Dec. 2023). DOI: [10.1007/s10456-023-09901-9](https://doi.org/10.1007/s10456-023-09901-9). URL: <https://inria.hal.science/hal-04358149>.
- [40] O. Solovyeva, M. Dimairo, C. Weir, S. W. Hee, A. Espinasse, M. Ursino, D. Patel, A. Kightley, S. Hughes, T. Jaki, A. Mander, T. Evans, S. Lee, S. Hopewell, K. R. Rantell, A.-W. Chan, A. Bedding, R. Stephens, D. Richards, L. Roberts, J. Kirkpatrick, J. de Bono and C. Yap. 'Development of consensus-driven SPIRIT and CONSORT extensions for early phase dose-finding trials: the DEFINE study'. In: *BMC Medicine* 21.1 (5th July 2023), p. 246. DOI: [10.1186/s12916-023-02937-0](https://doi.org/10.1186/s12916-023-02937-0). URL: <https://hal.science/hal-04336816>.
- [41] S. Tan, G. Goudot, A. Arnoux, Y. Tran, H. Maissoro, G. Poenou, G. Detriche, L. Khider, N. Mohamedi, T. Mirault, A. Galloula and E. Messas. 'Occurrence of Major Local Lower Limb Events in Type 2 Diabetic Patients with Lower Extremity Arterial Disease: Impact of Metformin'. In: *Annals of Vascular Surgery* 90 (Mar. 2023), pp. 153–161. DOI: [10.1016/j.avsg.2022.09.064](https://doi.org/10.1016/j.avsg.2022.09.064). URL: <https://hal.science/hal-04343656>.
- [42] R. Tsopra, N. Peiffer-Smadja, C. Charlier, F. Campeotto, C. Lemogne, P. Ruszniewski, B. Vivien and A. Burgun. 'Putting undergraduate medical students in AI-CDSS designers' shoes: An innovative teaching method to develop digital health critical thinking'. In: *International Journal of Medical Informatics* 171 (Mar. 2023), 10 p. DOI: [10.1016/j.ijmedinf.2022.104980](https://doi.org/10.1016/j.ijmedinf.2022.104980). URL: <https://hal.science/hal-04162799>.

- [43] S. Tzedakis, O. Yzhaky Shapira, M. Schwarzingler, S. Katsahian, A. Lazzati, A. Dohan, R. Coriat, P. Sogni, S. Pol, D. Fuks, V. Mallet, U. Marchese, S. Bouam and M. Alifano. 'COVID-19 and excess mortality of patients with liver cancer in France, January 2020–September 2022'. In: *British Journal of Surgery* (10th Nov. 2023). DOI: [10.1093/bjs/znad359](https://doi.org/10.1093/bjs/znad359). URL: <https://hal.science/hal-04348669>.
- [44] A. Vanier, J. Fernandez, S. Kelley, L. Alter, P. Semenzato, C. Alberti, S. Chevret, D. Costagliola, M. Cucherat, B. Falissard, F. Gueyffier, J. Lambert, E. Lengliné, C. Locher, F. Naudet, R. Porcher, R. Thiébaud, M. Vray, S. Zohar, P. Cochat and D. Le Guludec. 'Rapid access to innovative medicinal products while ensuring relevant health technology assessment. Position of the French National Authority for Health'. In: *BMJ Evidence-Based Medicine* (14th Feb. 2023), bmjebm-2022-112091. DOI: [10.1136/bmjebm-2022-112091](https://doi.org/10.1136/bmjebm-2022-112091). URL: <https://univ-rennes.hal.science/hal-04056273>.
- [45] A. Welfringer-Morin, P. Bataille, D. Drummond, N. Bellon, S. Ingen-Housz-Oro, J. Bonigen, S. Schmartz, L. Giraud-Kerleroux, F. Moulin, L. de Saint Blanquat, D. Bremond-Gignac, S. Hadj-Rabia and C. Bodemer. 'Comparison of idiopathic and drug-induced epidermal necrolysis in children'. In: *British Journal of Dermatology* 189.5 (25th Nov. 2023), pp. 631–633. DOI: [10.1093/bjd/ljad226](https://doi.org/10.1093/bjd/ljad226). URL: <https://hal.science/hal-04339226>.
- [46] C. Yap, J. Rekowski, M. Ursino, O. Solovyeva, D. Patel, M. Dimairo, C. Weir, A.-W. Chan, T. Jaki, A. Mander, T. Evans, R. Peck, K. Hayward, M. Calvert, K. R. Rantell, S. Lee, A. Kightley, S. Hopewell, D. Ashby, E. Garrett-Mayer, J. Isaacs, R. Golub, O. Kholmanskikh, D. Richards, O. Boix, J. Matcham, L. Seymour, S. P. Ivy, L. Marshall, A. Hommais, R. Liu, Y. Tanaka, J. Berlin, A. Espinasse and J. de Bono. 'Enhancing quality and impact of early phase dose-finding clinical trial protocols: SPIRIT Dose-finding Extension (SPIRIT-DEFINE) guidance'. In: *BMJ - British Medical Journal* (20th Oct. 2023), e076386. DOI: [10.1136/bmj-2023-076386](https://doi.org/10.1136/bmj-2023-076386). URL: <https://hal.science/hal-04336820>.
- [47] C. Yap, O. Solovyeva, J. de Bono, J. Rekowski, D. Patel, T. Jaki, A. Mander, T. Evans, R. Peck, K. Hayward, S. Hopewell, M. Ursino, K. R. Rantell, M. Calvert, S. Lee, A. Kightley, D. Ashby, A.-W. Chan, E. Garrett-Mayer, J. Isaacs, R. Golub, O. Kholmanskikh, D. Richards, O. Boix, J. Matcham, L. Seymour, S. P. Ivy, L. Marshall, A. Hommais, R. Liu, Y. Tanaka, J. Berlin, A. Espinasse, M. Dimairo and C. Weir. 'Enhancing reporting quality and impact of early phase dose-finding clinical trials: CONSORT Dose-finding Extension (CONSORT-DEFINE) guidance'. In: *BMJ - British Medical Journal* (20th Oct. 2023), e076387. DOI: [10.1136/bmj-2023-076387](https://doi.org/10.1136/bmj-2023-076387). URL: <https://hal.science/hal-04336824>.
- [48] C. Zamboglou, J. Peeken, A. Janbain, S. Katsahian, I. Strouthos, K. Ferentinos, A. Farolfi, S. Koerber, J. Debus, M. Vogel, S. Combs, A. Vrachimis, A. G. Morganti, S. Spohn, M. Shelan, D. Aebersold, A.-L. Grosu, F. Ceci, C. Henkenberens, S. Kroeze, M. Guckenberger, S. Fanti, C. Belka, P. Bartenstein, G. Hruby, S. Scharl, T. Wiegel, L. Emmett, A. Arnoux and N.-S. Schmidt-Hegemann. 'Development and Validation of a Multi-institutional Nomogram of Outcomes for PSMA-PET-Based Salvage Radiotherapy for Recurrent Prostate Cancer'. In: *JAMA Network Open* 6.5 (23rd May 2023), e2314748. DOI: [10.1001/jamanetworkopen.2023.14748](https://doi.org/10.1001/jamanetworkopen.2023.14748). URL: <https://hal.science/hal-04343654>.

International peer-reviewed conferences

- [49] L. Bleistein, A. Fermanian, A.-S. Jannot and A. Guilloux. 'Learning the Dynamics of Sparsely Observed Interacting Systems'. In: *PMLR - Proceedings of Machine Learning Research*. ICML 2023 - 40th International Conference on Machine Learning. Vol. 202. Honolulu, Hawaii, United States, 15th June 2023. URL: <https://hal.science/hal-04336559>.
- [50] L. Caquelin, P. Badra, M. Ursino, B. Laviolle and C. Locher. 'Meta-analysis of phase I dose-finding studies: Application for protein kinase inhibitors developed in oncology'. In: EPICLIN 2023 17e Conférence francophone d'Épidémiologie Clinique - 30e Journées des statisticiens des Centres de Lutte contre le Cancer. Fundamental and clinical pharmacology. Villers-lès-Nancy, France: Wiley, 2023. DOI: [10.1016/j.respe.2023.101816](https://doi.org/10.1016/j.respe.2023.101816). URL: <https://hal.science/hal-04266628>.
- [51] T. Fabacher, E.-A. Sauleau, N. Leclerc, H. Bergier, J.-E. Gottenberg, A. Coulet and A. Névoul. 'Evaluating the Portability of Rheumatoid Arthritis Phenotyping Algorithms: case study on French EHRs'. In: *Stud Health Technol Inform*. Medical Informatics Europe. 302. Gothenburg, Sweden, 18th May 2023, pp. 768–772. URL: <https://hal.science/hal-04069779>.

- [52] A. Neuraz, I. Lerner, O. Birot, C. Arias, L. Han, C. L. Bonzel, T. Cai, K. T. Huynh and A. Coulet. 'TAXN: Translate Align Extract Normalize, a multilingual extraction tool for clinical texts'. In: *MedInfo 2023 – the 19th world congress on Medical and Health Informatics*. Sydney, Australia, 8th July 2023. URL: <https://inria.hal.science/hal-04069590>.
- [53] M. A. N. Pour, A. Algergawy, P. Buche, L. J. Castro, J. Chen, A. Coulet, J. Cufi, H. Dong, O. Fallatah, D. Faria, I. Fundulaki, S. Hertling, Y. He, I. Horrocks, M. Huschka, L. Ibanescu, D. S. Jain, E. Jiménez-Ruiz, N. Karam, P. Lambrix, H. Li, Y. Li, P. Monnin, E. Nasr, H. Paulheim, C. Pesquita, T. Saveta, P. Shvaiko, G. Sousa, C. Trojahn, J. Vataščinová, M. Wu, B. Yaman, O. Zamazal and L. Zhou. 'Results of the Ontology Alignment Evaluation Initiative 2023'. In: *Proceedings of the 18th International Workshop on Ontology Matching co-located with the 22nd International Semantic Web Conference (ISWC 2023), Athens, Greece, November 7, 2023*. 18th International Workshop on Ontology Matching (OM 2023). Vol. 3591. CEUR workshop proceedings. Athens, Greece: CEUR-WS.org, 7th Nov. 2023, pp. 97–139. URL: <https://inria.hal.science/hal-04366893>.
- [54] K. Sedki, N. Boudegzame, J. Lamy and R. Tsopra. 'Learning Preferences in Lexicographic Choice Logic'. In: *Proceedings of the 15th International Conference on Agents and Artificial Intelligence - (Volume 3)*. ICAART 2023 - 15th International Conference on Agents and Artificial Intelligence. Lisbonne, Portugal: SCITEPRESS - Science and Technology Publications, 1st Jan. 2023, pp. 1012–1019. DOI: [10.5220/0011891300003393](https://hal.science/hal-04162803). URL: <https://hal.science/hal-04162803>.
- [55] F. Sverrisson, M. Akdel, D. Abramson, J. Feydy, A. Goncarenco, Y. Adeshina, D. Kovtun, C. Marquet, X. Zhang, D. Baugher, Z. Carpenter, L. Naef, M. M. Bronstein and B. Correia. 'DiffMaSIF: Surface-based Protein-Protein Docking with Diffusion Models'. In: *Machine Learning in Structural Biology workshop at NeurIPS 2023*. New Orleans, United States, 15th Dec. 2023. URL: <https://hal.science/hal-04360638>.

National peer-reviewed Conferences

- [56] N. Bannour, X. Tannier, B. Rance and A. Névéol. 'Positionnement temporel indépendant des événements : application à des textes cliniques en français'. In: *Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), volume 2 : travaux de recherche originaux – articles courts*. 18e Conférence en Recherche d'Information et Applications – 16e Rencontres Jeunes Chercheurs en RI – 30e Conférence sur le Traitement Automatique des Langues Naturelles – 25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues. Paris, France: ATALA, 2023, pp. 1–14. URL: <https://hal.science/hal-04130201>.

Scientific books

- [57] H. Pinnock, V. Poberezhets and D. Drummond. *Digital Respiratory Healthcare*. European Respiratory Society, 1st Dec. 2023. DOI: [10.1183/2312508X.erm10223](https://hal.science/hal-04339219). URL: <https://hal.science/hal-04339219>.

Scientific book chapters

- [58] S. Quennelle, M. Douillet, L. Friedlander, O. Boyer, A. Neuraz, A. Burgun and N. Garcelon. 'The Smart Data Extractor, a Clinician Friendly Solution to Accelerate and Improve the Data Collection During Clinical Trials'. In: *Caring is Sharing – Exploiting the Value in Data for Health and Innovation*. Vol. 302. Studies in Health Technology and Informatics. 18th May 2023. DOI: [10.3233/SHTI230112](https://hal.science/hal-04191423). URL: <https://hal.science/hal-04191423>.

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

- [59] A. Coulet, C. Gérardin, A. Névéol and X. Tannier, eds. *Proceedings of workshop on Patient Similarity, SimPa 2023*. 13th Mar. 2023. URL: <https://inria.hal.science/hal-04080808>.

Reports & preprints

- [60] C. Chadebec and S. Allasonnière. *Variational Inference for Longitudinal Data Using Normalizing Flows*. 13th Sept. 2023. URL: <https://hal.science/hal-04206103>.
- [61] J. Lambert, A.-L. Leutenegger, A. Baudot and A.-S. Jannot. *Improving patient clustering by incorporating structured label relationships in similarity measures*. 10th June 2023. DOI: [10.1101/2023.06.06.23291031](https://doi.org/10.1101/2023.06.06.23291031). URL: <https://hal.science/hal-04208019>.
- [62] A. Senellart, C. Chadebec and S. Allasonnière. *MultiVae: A Python library for Multimodal Generative Autoencoders*. 14th Sept. 2023. URL: <https://hal.science/hal-04207151>.
- [63] A. Senellart, C. Chadebec, S. Allasonnière and S. Allasonnière. *Improving Multimodal Joint Variational Autoencoders through Normalizing Flows and Correlation Analysis*. 16th May 2023. URL: <https://hal.science/hal-04207134>.