# Internship proposal

# Federated Causal Discovery for Understanding PrEP Uptake and Sexual Behavior Change among MSM in France and Portugal

**Context:** Understanding the relations between pre-exposure prophylaxis (PrEP) uptake and sexual behavior change among men who have sex with men (MSM) is becoming increasingly important for guiding HIV prevention strategies. Estimating causal effects in this context requires a causal graph [Pearl, 2000, Hernan and Robins, 2023], which serves as a guide for identifying the variables that must be adjusted for. However, specifying such a graph solely from expert knowledge is often challenging. This makes causal discovery [Spirtes et al., 2000] (the process of learning causal graph directly from data) a promising approach. However, applying causal discovery within a single dataset, faces several limitations. For example, datasets collected in one country may be affected by specific interventions or limited sample sizes. Leveraging data from multiple sources (e.g., different countries) can help overcome these limitations by providing complementary information that allows for better orientation of causal relations and improved robustness.

In recent years, the availability of rich behavioral, clinical, and survey data describing MSM populations has increased across multiple countries. Yet, strict data protection and privacy regulations (e.g., GDPR) prevent the pooling of individual-level data across sites. This limits our ability to perform cross-country causal analyses that could reveal how social, behavioral, and contextual factors interact to influence PrEP uptake and sexual practices.

Recently, a new federated causal discovery algorithm called I-PERI [Baldo and Assaad, 2025] has been proposed. I-PERI allows multiple data sites to collaboratively infer a causal graph without exchanging raw data, while accounting for potential site-specific interventions. However, this algorithm was developed in a purely theoretical context and it has not yet been tested or adapted for real-world data. Moreover, the original algorithm assumes ideal conditions (no unmeasured confounding, identical variable sets, perfect measurements, and no missing data), which rarely hold in epidemiological research.

**Proposal:** This internship aims to adapt and apply the I-PERI algorithm to discover causal relations between variables related to PrEP uptake and sexual behavior change among MSM, using two complementary datasets: one collected in France and one in Portugal. Specifically, the intern will:

- Explore the two datasets and consult domain experts to determine which assumptions of the I-PERI algorithm are not satisfied.
- Adapt the I-PERI algorithm to handle realistic data conditions while ensuring privacy preservation under a federated setup.
- Apply the adapted algorithm to the French and Portuguese datasets to infer a shared causal graph describing how behavioral, clinical, and social factors jointly influence PrEP uptake and sexual behavior change.

**Required skills:** Highly motivated candidate with an M2 degree and strong background in probability, machine learning, and causal inference, along with a keen interest in epidemiology. Proficiency in programming is also required.

**Location:** The intern will work at IPLESP (`https://iplesp.fr/`), located in Paris. She/he will be supervised by Federico Baldo, Eugenio Valdano, and Charles Assaad.

**Dates:** Starting date: To be discussed, early 2026, for a duration of 5-6 months.

**Contact:** To apply, please send a CV and a cover letter to Federico Baldo federico.baldo@inserm.fr

## References

F. Baldo and C. K. Assaad. Federated causal discovery with uknown interventions, 2025.

M. Hernan and J. Robins. *Causal Inference: What If*. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. CRC Press, 2023. ISBN 9781420076165.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000. ISBN 0-521-77362-8.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT press, 2nd edition, 2000.