A Bayesian Pharmacometric Machine Learning Framework for Integrating Historical Controls in Drug Development

**PhD Proposal**

Marothi Peter Letsoalo

Table of contents

## Introduction

The development of new drugs represents a multifaceted challenge, marked by complexity, significant costs, and pressing ethical considerations, with clinical trials serving as the pivotal mechanism for evaluating drug efficacy and safety (Friedman, Furberg, and DeMets 2010). Traditionally, these trials necessitate control groups that receive either a placebo or standard treatment, which can pose substantial ethical and recruitment hurdles, especially in scenarios where effective therapies are already in use. The reliance on concurrent control groups in randomized controlled trials (RCTs) often results in prolonged drug development periods, inflated costs, and ethical quandaries, particularly when it involves withholding potential treatments from patients (Pocock 2013).

In response to these challenges, the utilization of historical control data (HCD) from prior studies has emerged as a significant area of interest (Berry and Hardwick 1994; Marion and Althouse 2023). HCD holds the promise of diminishing the need for placebo or standard treatment arms, thus reducing patient burden and expediting drug development timelines (Viele et al. 2013; Lim et al. 2019). By integrating data from previous trials, researchers can make more informed decisions regarding the effectiveness and safety of new interventions, potentially facilitating quicker approvals and access to novel treatments (Hall et al. 2020). This approach is especially valuable in the context of rare diseases or small patient populations, where assembling a sufficient number of participants for a traditional RCT may prove challenging (Ghadessi et al. 2020).

However, the intrinsic heterogeneity and potential biases in HCD, such as variations in patient demographics, study designs, outcome measures, and potential publication or selection biases, pose substantial challenges to their straightforward integration (Schmidli et al. 2014; Branders et al. 2021). Discrepancies in inclusion/exclusion criteria, disease severity, concurrent medications, and data collection methods can all contribute to this heterogeneity, making direct comparisons arduous (Hall et al. 2020). Moreover, the quality of HCD can vary significantly, with some studies lacking adequate documentation or comprehensive reporting, further complicating the integration process.

To navigate these complexities, a synergistic approach combining Bayesian statistics, pharmacometric modeling, and machine learning is proposed. Bayesian methods provide a flexible framework that facilitates the integration of prior information, such as HCD, allowing for continuous updates in treatment effect beliefs as new data becomes available (Jin et al. 2023). Pharmacometric models, which are adept at capturing the dynamic interplay between drug action and disease progression, can be integrated with Bayesian methods to account for inter-individual variability and predict treatment outcomes (Wakefield 1996; Lunn 2005). Additionally, machine learning algorithms can enhance the selection and validation of relevant HCD by detecting patterns and relationships within data, thus improving the predictive accuracy of models (Lee et al. 2020).

This proposed research aims to unify these methodologies into a comprehensive framework for integrating HCD in drug development. By addressing challenges related to data selection, weighting, model integration, and regulatory alignment, this initiative strives to develop and validate a robust tool for incorporating HCD, ultimately contributing to more efficient and ethically sound drug development processes.

## Rationale

The rationale for this research is rooted in the critical need for a robust and efficient method to integrate historical control data (HCD) in drug development. By harnessing the capabilities of Bayesian statistics, pharmacometric modeling, and machine learning, this research aims to address the inherent challenges of heterogeneity and biases within HCD. Successful development and validation of this framework could significantly accelerate drug development timelines, reduce patient burden, and improve the ethical conduct of clinical trials. Moreover, this research promises to enhance decision-making processes in drug development, ultimately enabling quicker access to safe and effective treatments for patients.

## Aim and Objectives

To develop and validate a Bayesian pharmacometric machine learning framework that integrates historical control data in drug development, thereby improving efficiency, ethical standards, and decision-making in clinical trials.

To achieve this overarching aim, the following specific objectives will be pursued:

1. Construct a unified framework combining Bayesian statistics, pharmacometric modeling, and machine learning algorithms to facilitate the quantification of uncertainty, predict treatment effects using HCD, and identify patterns and relationships within the data.
2. Design and implement algorithms to accurately identify and select relevant HCD based on their similarity to the current trial population, while accounting for potential biases and heterogeneity.
3. Incorporate pharmacokinetic/pharmacodynamic (PK/PD) modeling into the framework to predict treatment effects in the current trial population accurately, adjusting for patient characteristics and disease progression.

### Additional Objectives

1. Investigate the framework’s potential to enable adaptive trial design adjustments, focusing on utilizing accumulating HCD to potentially reduce sample size requirements while maintaining statistical rigor and ethical considerations.
2. Examine the regulatory implications of using HCD in clinical trials to ensure compliance with guidelines from agencies such as the FDA and EMA. Validate the framework thoroughly to ensure its reliability and robustness for decision-making.

## Methodology

The proposed research will follow a multi-phase approach, ensuring the robust and reliable implementation of the Bayesian pharmacometric machine learning (BPML) framework for integrating historical control data in clinical trials.

### Phase 1: Framework Construction

* Begin with a comprehensive literature review to identify existing methodologies and pinpoint gaps in integrating HCD in drug development.
* Develop the foundation of the framework using Bayesian statistical methods to incorporate prior information and quantify uncertainty.
* Integrate pharmacometric models to capture the complex dynamics of drug effects and disease progression.
* Employ machine learning algorithms to detect patterns and predict treatment effects based on HCD.

### Phase 2: Algorithm Design

* Develop algorithms for identifying and selecting relevant HCD, using similarity metrics and advanced machine learning techniques to manage heterogeneity and biases.
* Design methods for optimally weighting HCD, employing Bayesian hierarchical models to balance the contributions of various data sources.

### Phase 3: PK/PD Model Integration

* Integrate PK/PD modeling within the framework following the identification and selection of HCD, adjusting models to reflect the characteristics of the current trial population.
* Use Bayesian methods to update models with new data, providing contextually relevant predictions of treatment effects.

### Phase 4: Framework Validation

* Conduct simulations and real-world case studies to validate the framework, testing its robustness under various scenarios and providing practical insights into its performance and applicability.

### Additional Phases

#### Phase 5: Adaptive Trial Design Exploration

* Explore the framework’s capability to utilize accumulating HCD during ongoing trials to inform interim analyses and potentially reduce sample size requirements.

#### Phase 6: Regulatory Compliance

* Investigate the regulatory implications of using HCD, ensuring the framework’s compliance with guidelines and its reliability for decision-making.

## Ethical Considerations

Throughout the research, ethical considerations will be paramount. The integration of HCD aims to enhance the ethical conduct of clinical trials by potentially reducing the need for control groups and minimizing patient exposure to less effective treatments. The methodology will be designed to ensure transparency, reproducibility, and adherence to regulatory standards (Taylor, Rebok, and Marsiske 2022).

## References

Aim and Objectives The overarching aim of this research is to develop and validate a Bayesian pharmacometric machine learning framework for the integration of historical control data (HCD) in drug development, thereby enhancing the efficiency, ethical conduct, and decision-making processes within clinical trials.

To achieve this overarching aim, the following specific objectives will be pursued:

Framework Development: Construct a unified Bayesian pharmacometric machine learning (BPML) framework that integrates Bayesian statistics and machine learning algorithms to facilitate the effective integration of HCD in clinical trials. Algorithm Design for HCD Selection and Weighting: Develop and implement algorithms to identify and select relevant HCD based on their similarity to the current trial population, while accounting for potential biases and heterogeneity. Design methods to optimally weight HCD, ensuring that the most pertinent data influence the analysis. Verification of HCD Comparability Using Machanistic Pharmacometric Models: Integrate pharmacometric models to verify the comparability of selected HCD with the current trial data ensuring that HCD accurately reflects disease dynamics and biological Machanistic observed in the current trial population. Biostatistical Model Integration for Treatment Effect Prediction: Integrate common biostatistical models into the framework to evaluate the efficacy and safety of treatments based on the verified HCD. The objective is to ensure the selected HCD and the trial cohort are providing a comprehensive assessment of treatment effects as expected using robust biostatistical techniques. Adaptive Trial Design Exploration: Investigate the potential of the BPML framework to enable adaptive trial design adjustments, utilizing accumulating HCD to potentially reduce sample size requirements while maintaining statistical rigor and ethical considerations.

Berry, Donald A., and Janis P. Hardwick. 1994. “Using Historical Controls in Clinical Trials: Application to ECMO.” In, 141–56. Springer New York. <https://doi.org/10.1007/978-1-4612-2618-5_12>.

Branders, Samuel, Alvaro Pereira, Guillaume Bernard, Marie Ernst, Jamie Dananberg, and Adelin Albert. 2021. “Leveraging Historical Data to Optimize the Number of Covariates and Their Explained Variance in the Analysis of Randomized Clinical Trials.” *Statistical Methods in Medical Research* 31 (2): 240–52. <https://doi.org/10.1177/09622802211065246>.

Friedman, Lawrence M., Curt D. Furberg, and David L. DeMets. 2010. *Fundamentals of Clinical Trials*. Springer New York. <https://doi.org/10.1007/978-1-4419-1586-3>.

Ghadessi, Mercedeh, Rui Tang, Joey Zhou, Rong Liu, Chenkun Wang, Kiichiro Toyoizumi, Chaoqun Mei, Lixia Zhang, C. Q. Deng, and Robert A. Beckman. 2020. “A Roadmap to Using Historical Controls in Clinical Trials by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG).” *Orphanet Journal of Rare Diseases* 15 (1). <https://doi.org/10.1186/s13023-020-1332-x>.

Hall, Kathryn T., Lene Vase, Deirdre K. Tobias, Hesam T. Dashti, Jan Vollert, Ted J. Kaptchuk, and Nancy R. Cook. 2020. “Historical Controls in Randomized Clinical Trials: Opportunities and Challenges.” *Clinical Pharmacology & Therapeutics* 109 (2): 343–51. <https://doi.org/10.1002/cpt.1970>.

Jin, Huaqing, Mi-Ok Kim, Aaron Scheffler, and Fei Jiang. 2023. “Bayesian Adaptive Design for Covariate-Adaptive Historical Control Information Borrowing.” *Statistics in Medicine* 42 (29): 5338–52. <https://doi.org/10.1002/sim.9913>.

Lee, Jeong-Hee, Jongseok Kang, We Shim, Hyun-Sang Chung, and Tae-Eung Sung. 2020. “Pattern Detection Model Using a Deep Learning Algorithm for Power Data Analysis in Abnormal Conditions.” *Electronics* 9 (7): 1140. <https://doi.org/10.3390/electronics9071140>.

Lim, Jessica, Li Wang, Nicky Best, Jeen Liu, Jiacheng Yuan, Florence Yong, Lanju Zhang, et al. 2019. “Reducing Patient Burden in Clinical Trials Through the Use of Historical Controls: Appropriate Selection of Historical Data to Minimize Risk of Bias.” *Therapeutic Innovation & Regulatory Science* 54 (4): 850–60. <https://doi.org/10.1007/s43441-019-00014-4>.

Lunn, David J. 2005. “Bayesian Analysis of Population Pharmacokinetic/Pharmacodynamic Models.” In, 351–70. Springer-Verlag. <https://doi.org/10.1007/1-84628-119-9_11>.

Marion, Joe D., and Andrew D. Althouse. 2023. “The Use of Historical Controls in Clinical Trials.” *JAMA* 330 (15): 1484. <https://doi.org/10.1001/jama.2023.16182>.

Pocock, Stuart J. 2013. “Clinical Trials,” July. <https://doi.org/10.1002/9781118793916>.

Schmidli, Heinz, Sandro Gsteiger, Satrajit Roychoudhury, Anthony O’Hagan, David Spiegelhalter, and Beat Neuenschwander. 2014. “Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information.” *Biometrics* 70 (4): 1023–32. <https://doi.org/10.1111/biom.12242>.

Taylor, Brad P., George W. Rebok, and Michael Marsiske. 2022. “Good Clinical Practice Improves Rigor and Transparency: Lessons from the ACTIVE Trial.” *Psychology and Aging* 37 (1): 43–50. <https://doi.org/10.1037/pag0000653>.

Viele, Kert, Scott Berry, Beat Neuenschwander, Billy Amzal, Fang Chen, Nathan Enas, Brian Hobbs, et al. 2013. “Use of Historical Control Data for Assessing Treatment Effects in Clinical Trials.” *Pharmaceutical Statistics* 13 (1): 41–54. <https://doi.org/10.1002/pst.1589>.

Wakefield, Jon. 1996. “The Bayesian Analysis of Population Pharmacokinetic Models.” *Journal of the American Statistical Association* 91 (433): 62–75. <https://doi.org/10.1080/01621459.1996.10476664>.