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 is a complex, costly, and ethically challenging endeavor, with clinical trials playing a critical role in assessing drug efficacy and safety (Friedman, Furberg, and DeMets 2010) . Traditionally, these trials often involve control groups receiving placebo or standard treatment, raising ethical concerns and recruitment challenges, particularly when effective therapies already exist . The reliance on concurrent control groups in randomized controlled trials (RCTs) can lead to delays in drug development, increased costs, and potential ethical dilemmas when withholding potentially effective treatments from patients (Pocock 2013).

To address these issues, leveraging historical control data (HCD) from previous studies has gained significant interest (Berry and Hardwick 1994; Marion and Althouse 2023). HCD offers the potential to reduce the need for placebo or standard treatment arms, thereby minimizing patient burden and accelerating drug development timelines (Viele et al. 2013; Lim et al. 2019). By incorporating data from past trials, researchers can make more informed decisions about the effectiveness and safety of new interventions, potentially leading to faster approval and access to new treatments (Hall et al. 2020). The use of HCD can be particularly valuable in rare diseases or small patient populations where recruiting a sufficient number of participants for a traditional RCT may be difficult (Ghadessi et al. 2020).

However, the heterogeneity and potential biases inherent in HCD, such as variations in patient populations, study designs, outcome measures, and potential publication or selection biases, present significant obstacles to their seamless integration (Schmidli et al. 2014; Branders et al. 2021). Differences in inclusion/exclusion criteria, disease severity, concomitant medications, and data collection methods can all contribute to heterogeneity, making direct comparisons challenging (Hall et al. 2020). Additionally, the quality of HCD can vary considerably, with some studies lacking adequate documentation or reporting, further complicating the integration process.

Bayesian statistics, pharmacometric modeling, and machine learning offer a synergistic approach to address these challenges. Bayesian methods provide a flexible framework for incorporating prior information, such as HCD, into current analyses, allowing for the updating of beliefs about treatment effects as new data becomes available (Jin et al. 2023). Pharmacometric models can be integrated with Bayesian methods to capture complex dose-response relationships, account for inter-individual variability, and predict treatment outcomes (Wakefield 1996; Lunn 2005). Machine learning algorithms can aid in identifying relevant HCD, given their ability to detect patterns and relationships within data, and improving the predictive accuracy of models (Lee et al. 2020).

The proposed framework aims to synergize these approaches to create a comprehensive solution for integrating HCD in drug development. By addressing challenges related to data selection, weighting, model integration, and regulatory alignment, this research aims to develop and validate a robust tool for integrating HCD, ultimately contributing to more efficient and ethical drug development processes.

## Rationale

The rationale for this research is to address the critical need for a robust and efficient framework for integrating historical control data (HCD) in drug development. By leveraging Bayesian statistics, pharmacometric modeling, and machine learning, this research aims to overcome the challenges associated with HCD heterogeneity and potential biases. The successful development and validation of such a framework would have significant implications for accelerating drug development timelines, reducing patient burden, and improving the ethical conduct of clinical trials. Furthermore, this research could contribute to more informed decision-making in drug development, ultimately leading to faster access to safe and effective therapies for patients in need.

## 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:

1. A unified framework will be constructed, integrating Bayesian statistics, pharmacometric modeling, and machine learning algorithms. This framework will facilitate the quantification of uncertainty, the prediction of treatment effects based on HCD, and the identification of patterns and relationships within the HCD.
2. Algorithms will be designed and implemented to identify and select relevant HCD based on their similarity to the current trial population, while accounting for potential biases and heterogeneity. Furthermore, methods will be developed to optimally weight HCD based on their relevance and informativeness, ensuring that the most pertinent data influence the analysis.
3. Pharmacokinetic/pharmacodynamic (PK/PD) modeling will be incorporated into the framework to predict treatment effects in the current trial population based on HCD. This integration will involve adjusting for differences in patient characteristics and disease progression to ensure accurate and contextually relevant predictions.

### Additional Objectives

1. The potential of the proposed framework to enable adaptive trial design adjustments will be investigated. This exploration will focus on utilizing accumulating HCD to potentially reduce sample size requirements while maintaining statistical rigor and ethical considerations.
2. The regulatory implications of using HCD in clinical trials will be thoroughly investigated to ensure compliance with guidelines from agencies such as the FDA and EMA. The framework will undergo rigorous validation to ensure its reliability and robustness for decision-making, providing a solid foundation for regulatory acceptance and practical implementation.

## Methodology

The proposed research will be conducted through a multi-phase approach, encompassing the development, validation, and application of a Bayesian pharmacometric machine learning (BPML) framework for integrating historical control data (HCD) in clinical trials. This methodology is designed to ensure the robust and reliable implementation of the framework, which synergistically combines Bayesian statistics, pharmacometric modeling, and machine learning algorithms.

### Phase 1: Framework Construction

The initial phase involves the construction of the unified BPML framework. This will commence with a comprehensive literature review to identify existing methodologies and potential gaps in the integration of HCD in drug development (e.g., Viele et al. (2013)). Bayesian statistical methods will form the foundation of the framework, allowing for the incorporation of prior information and the quantification of uncertainty (Schmidli et al. 2014). Pharmacometric models, specifically pharmacokinetic/pharmacodynamic (PK/PD) models, will be integrated to capture the complex dynamics of drug effects and disease progression (Mould and Upton 2013). Machine learning algorithms will be employed to identify patterns and relationships within the HCD, facilitating the prediction of treatment effects (Vamathevan et al. 2019).

### Phase 2: Algorithm Design

In this phase, algorithms will be designed to identify and select relevant HCD. This involves the development of similarity metrics to match historical controls to the current trial population, accounting for demographic and clinical characteristics, disease states, and other pertinent variables. Advanced machine learning techniques, such as clustering and classification algorithms, will be utilized to manage the heterogeneity and potential biases inherent in historical data (Schmidli et al. 2014). Methods for optimally weighting HCD will be developed, ensuring that the analysis is influenced by the most relevant and informative data. This weighting process will leverage Bayesian hierarchical models to balance the contributions of various data sources (Schmidli et al. 2014).

### Phase 3: PK/PD Model Integration

The integration of PK/PD modeling within the framework will follow the identification and selection of HCD. This step involves adjusting the pharmacometric models to reflect the characteristics of the current trial population. Bayesian methods will be used to update these models with new data, providing contextually relevant predictions of treatment effects (Jin et al. 2023). The models will be calibrated using a combination of historical and current trial data, with careful attention to differences in patient characteristics and disease progression (Mould and Upton 2013).

### Phase 4: Framework Validation

Validation of the framework will be conducted through a series of simulations and real-world case studies. Simulations will be designed to test the robustness of the framework under various scenarios, including different levels of data heterogeneity, varying sample sizes, and different degrees of similarity between historical and current trial populations. These simulations will help identify potential limitations and areas for improvement. Real-world case studies, involving actual clinical trial data, will provide practical insights into the framework’s performance and applicability. Metrics for validation will include predictive accuracy, the precision of uncertainty estimates, and the framework’s ability to improve decision-making processes in clinical trials.

### Additional Phases

#### Phase 5: Adaptive Trial Design Exploration

The framework will be examined for its capability to utilize accumulating HCD during ongoing trials, exploring the potential of adaptive trial design adjustments. This exploration will assess how real-time integration of HCD can inform interim analyses and potentially reduce sample size requirements while maintaining statistical rigor and ethical considerations. The framework’s flexibility and adaptability will be tested through simulations that mimic adaptive trial scenarios (Pallmann et al. 2018).

#### Phase 6: Regulatory Compliance

The regulatory implications of using HCD in clinical trials will be thoroughly investigated to ensure compliance with guidelines from regulatory agencies such as the FDA and EMA (Beckett 2014; EMA 2014). The framework will undergo rigorous validation to ensure its reliability and robustness for decision-making, providing a solid foundation for regulatory acceptance and practical implementation.

## 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

Beckett, R. D. 2014. “Food and Drug Administration, US.” In, 631–32. Elsevier. <https://doi.org/10.1016/b978-0-12-386454-3.00314-6>.

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>.

EMA. 2014. “The European Medicines Agency (EMA).” In, 557. Elsevier. <https://doi.org/10.1016/b978-0-12-386454-3.00564-9>.

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>.

Mould, DR, and RN Upton. 2013. “Basic Concepts in Population Modeling, Simulation, and Model-Based Drug DevelopmentPart 2: Introduction to Pharmacokinetic Modeling Methods.” *CPT: Pharmacometrics & Systems Pharmacology* 2 (4): 1–14. <https://doi.org/10.1038/psp.2013.14>.

Pallmann, Philip, Alun W. Bedding, Babak Choodari-Oskooei, Munyaradzi Dimairo, Laura Flight, Lisa V. Hampson, Jane Holmes, et al. 2018. “Adaptive Designs in Clinical Trials: Why Use Them, and How to Run and Report Them.” *BMC Medicine* 16 (1). <https://doi.org/10.1186/s12916-018-1017-7>.

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>.

Vamathevan, Jessica, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, et al. 2019. “Applications of Machine Learning in Drug Discovery and Development.” *Nature Reviews Drug Discovery* 18 (6): 463–77. <https://doi.org/10.1038/s41573-019-0024-5>.

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>.