Project Details	
Project Code	MRCPHS26Br Millard
Title	Predicting adverse pregnancy outcomes using statistical and machine learning approaches
Research Theme	PHS
Project Type	Dry lab
Summary	There are multiple adverse events or health outcomes that can occur during pregnancy or shortly after birth such as emergency caesarean section, pre-term birth, and complications meaning that the newborn is admitted to the neonatal intensive care unit. Predicting these adverse events is essential for identifying those at high risk so that monitoring or more targeted care can be provided. This project aims to develop and evaluate statistical and machine learning models for predicting adverse events during pregnancy, in a way which is explainable to the clinicians and patients.
Description	**Background:** There are multiple adverse events or health outcomes that can occur during pregnancy or shortly after birth, including the pregnant individual having a miscarriage, pre-term birth, or requiring an emergency caesarean section, and complications meaning that the newborn is admitted to the neonatal intensive care unit (NICU). Predicting risk of these events is essential to provide additional monitoring and targeted care for those at higher risk. Electronic health records (EHR) collected during pregnancy include a wealth of information on the pregnant individual including their medical and obstetric history and measurements recorded throughout pregnancy (e.g. blood pressure and blood tests), and information on the developing foetus/newborn such as growth scans and measures recorded at birth (e.g. birth weight). There is much potential to use these data for predicting adverse events during pregnancy to improve outcomes. For instance, a recent review of prediction models for emergency caesarean section [1] identified an urgent need for a prediction model that performs well across diverse clinical backgrounds. Few studies have been conducted evaluating prediction models for admission to NICU, and those that have were in small samples, lacked external validation or considered a limited number of predictor variables [2-4]. **Aim:** The aim of this PhD is to evaluate modelling approaches for predicting adverse pregnancy outcomes and develop a deployable model (or models) that can be used by obstetricians to improve patient care. Data: This project will use EHR data from Bristol and Liverpool (~60K patients each). **Objectives:** The PhD will start with a review of the current literature and the student will define objectives depending on previous research, and their own interests and skills development aims. Possible objectives: - Evaluate approaches for predicting time-to-event outcomes, including flexible parametric models and machine learning time-to-event models

	such as random survival forests [5] and deep survival machines [6], and potentially with self-supervised learning [7-8]. - Compare different approaches for modelling time varying features, where a measure such as blood pressure is taken several times during pregnancy and the change over time may be predictive of adverse events [9]. - Explore modelling approaches that simultaneously predict multiple adverse outcomes (e.g. transformer deep learning approaches [10]) or allow for competing risks [6]. - Assess the trade-off between model performance and model size, both in terms of the complexity of the model and number of predictor variables included. For example, can a smaller model be developed with similar performance to a large 'state-of-the-art' machine learning model, that has a lower carbon footprint and is easier to deploy across diverse settings (as available data might differ across health care settings). - Assessing approaches for appropriately evaluating performance of prediction models depending on the specific objectives of the prediction task, which could include a review of the common issues in studies evaluating perinatal prediction models (including the quality of reporting of these studies). - Exploring how to explain the prediction models to better serve clinicians and patients, e.g., in the form of natural language using Large Language Models [11]. Evaluation of prediction models will include both internal validation (evaluation on the Bristol dataset used for developing the models e.g. using cross-validation), and external validation (evaluation using the Liverpool data that will not be used during model development), and further with clinical experts' validation of how the model is explained. **References:** [1] A. Hunt et al. (2025). Systematic Review of Clinical Prediction Models for the Risk of Emergency Caesarean Births. BJOG, 132:231-240. [2] G. Marvin et al. (2021). "Explainable Feature Learning for Predicting Neonatal Intensive Care Unit (NICU) Admissions," 2021 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), Dhaka, Bangladesh, 69-74. [3] R. Tashakkori et al.(2024). The prediction of NICU admission and identifying influential factors in four different categories leveraging machine learning approaches. Biomedical Signal Processing and Control 90. 105844. [4] M Mahendra et al. (2020). Predicting NICU admissions in near-term and term infants with low illness acuity. J Perinatol. 41(3):478-485. [5] H Ishwaran et al. (2008). Random survival forests. Ann. Appl. Stat. 2(3):841-860. [6] C. Nagpal et al. (2021). Deep Survival Machines: Fully Parametric Survival Regression and Representation Learning for Censored Data With Competing Risks, in IEEE Journal of Biomedical and Health Informatics, 25(8):3163-3175. [7] R. Krishnan et al. (2022). Self-supervised learning in medicine and healthcare. Nat. Biomed. Eng. 6:1346–1352.

	[8] E. Steinberg et al. (2023). MOTOR: A Time-to-Event Foundation Model For Structured Medical Records. The Twelfth International Conference on Learning Representations. [9] A. Cascarano et al. (2023). Machine and deep learning for longitudinal biomedical data: a review of methods and applications. Artif Intell Rev. 56(Suppl 2):1711–1771. [10] L. Rasmy et al. (2021). Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. npj Digit. Med. 4:86. [11] A. Sivaprasad et al. (2025). Patient-Centred Explainability in IVF Outcome Prediction. (No. arXiv:2506.18760). arXiv.
Supervisory Team	
Lead Supervisor	
Name	Dr Louise Millard
Affiliation	Bristol
College/Faculty	Faculty of Health and Life Sciences
Department/School	Population Health Sciences, Bristol Medical School
Email Address	louise.millard@bristol.ac.uk
Co-Supervisor 1	
Name	Dr Abi Merriel
Affiliation	Other
College/Faculty	University of Liverpool
Department/School	Faculty of Health and Life Sciences
Co-Supervisor 2	
Name	Dr Hang Dong
Affiliation	Exeter
College/Faculty	Faculty of Environment, Science and Economy
Department/School	Department of Computer Science
Co-Supervisor 3	
Name	Dr Gemma Clayton
Affiliation	Bristol
College/Faculty	Faculty of Health and Life Sciences
Department/School	Population Health Sciences, Bristol Medical School