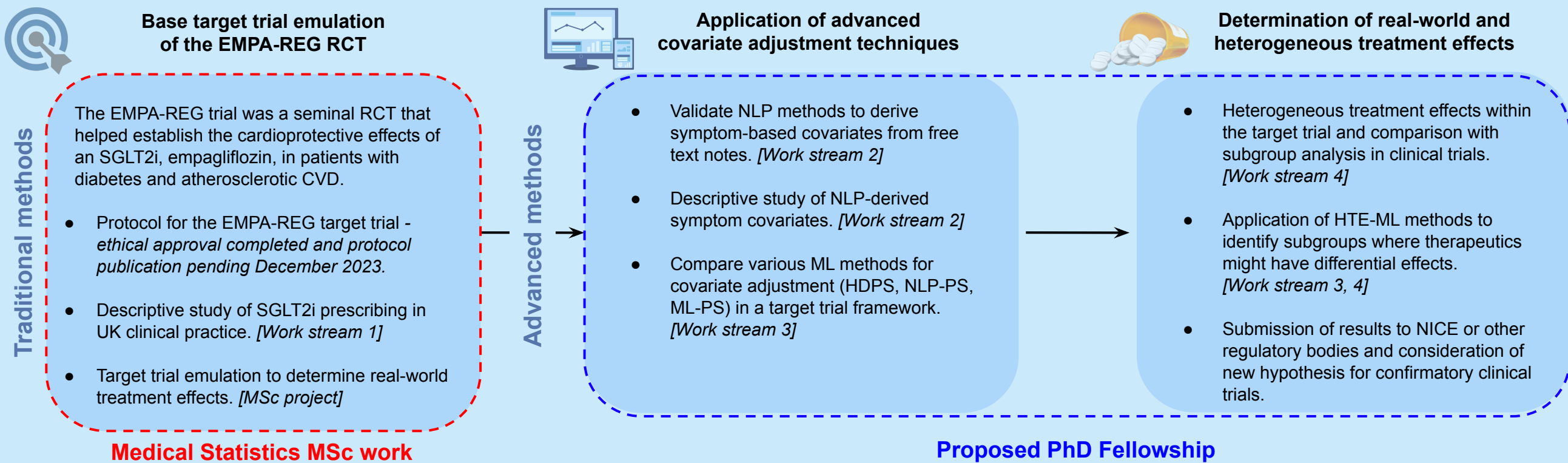


Objectives: Applying new machine learning and causal inference techniques for a better understanding of real-world drug effects



- 1: Compare machine learning and traditional methodologies in pharmacoepidemiology research → Promote the **robust** analysis of real-world data.
- 2: Gain a better understanding of real-world treatment effects of anti-diabetes therapeutics → Leverage real-world data to **better inform patients and clinicians**.
- 3: Study real-world heterogeneous treatment effects for anti-diabetes therapeutics → Develop evidence to **inform new hypotheses**.

Study design: Machine learning and causal inference methods will be applied to a baseline target trial emulation of an important cardiovascular outcome trial



PhD Impact:



- Patients and clinicians:**
- What is the real effect of taking this drug?*
- Drug regulators, guideline committees and pharmaceutical industry:**
- What gaps in evidence can this data address?
 - What new hypotheses can be generated from real-world evidence?
- Statisticians and Epidemiologists:**
- What is the optimal way to analyse real-world clinical data?
 - How can ML techniques reduce residual confounding?

Timeline:



- Year 1: Preparation**
- Publication of TTE [work stream 1]
 - Validate NLP methods [work stream 2]
 - Descriptive NLP work [work stream 2]
- Year 2: Propensity score**
- NLP-PS
 - HDPS
 - ML-PS
- Year 3: PPI and treatment effect heterogeneity**
- PPI panels
 - HTE work - causal forests and meta-learners

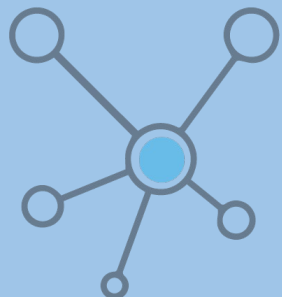


Figure 1: Hypotheses and study design for proposed PhD. SGLT2i: sodium-glucose cotransporter-2 inhibitor; RCT: randomised controlled trial; RW: real-world; NLP: natural language processing; TTE: target trial emulation; ML: machine learning; HTE: heterogeneous treatment effects; NICE: National institute for health and care excellence; PS: propensity score; HDPS: high-dimensionality PS; PPI: patient and public involvement.