Research Proposal:
Machine Learning Approaches to Assess Long-term Inclisiran Safety

Slide 1 – Title

Hello, my name is Guilherme, and welcome to my PgDip research proposal presentation. The project I am presenting today is entitled: *Machine Learning Approaches to Assess Long-term Inclisiran Safety*.
This proposal lies at the intersection of artificial intelligence, pharmacovigilance, and clinical trial research.
My aim is to show how machine learning can be applied to a large, ongoing clinical trial to answer an important and yet unresolved question: what are the long-term safety implications of inclisiran, a novel small interfering RNA therapy?

Slide 2 – Significance of the Research

Adverse drug events, often shortened to ADEs, are a leading cause of patient harm globally. They contribute not only to morbidity and mortality, but also to increased healthcare utilisation, and economic burden (Kim *et al*., 2022).

Although clinical trials are carefully designed to assess both efficacy and safety, they are often underpowered to detect rare, long-term, or unexpected harms. This is particularly true for novel classes of therapy where biological mechanisms differ from traditional drugs.

Cardiovascular disease is the leading cause of death worldwide, accounting for 17.9 million deaths every year (World Health Organization, 2023). Statins remain the cornerstone of lipid-lowering therapy, but residual cardiovascular risk persists. This has driven development of new therapeutic options, including PCSK9-targeting therapies.

Inclisiran is one of these new therapies: a small interfering RNA drug that silences PCSK9 gene expression in the liver. It leads to sustained reductions of LDL cholesterol in the range of 44 to 54 per cent, and is uniquely convenient because it is dosed only twice per year as an injection (Marrs and Anderson, 2024; Ray *et al*., 2020). This infrequent dosing distinguishes it from PCSK9 monoclonal antibodies, which require more frequent injections, and it raises both opportunities for adherence and questions about long-term safety.

The currently available trial evidence suggests that inclisiran is generally safe and well tolerated, with injection-site reactions being the most consistent adverse effect (Wright *et al*., 2023). However, long-term data are sparse, with most published evidence extending to about six years at most. Importantly, the full spectrum of potential long-term harms, especially those that may be subtle or emerge only after years of exposure, is not yet known.

This is where the ORION-4 trial becomes particularly valuable. ORION-4 is a large cardiovascular outcomes trial with 15,000 participants who all have established atherosclerotic cardiovascular disease. The trial is expected to run until at least 2026, with extended follow-up and national data linkage planned for the 12,000 participants recruited in the UK (ClinicalTrials.gov, 2023). As such, it represents a unique opportunity to assess both efficacy and long-term safety of inclisiran in a large, diverse patient population. Moreover, the randomised controlled trial design facilitates robust causal inference, which can be challenging to appropriately undertake in observational

analyses subject to biases and subtle differences between patients that receive the drug and those that do not.

My proposal takes advantage of this opportunity by suggesting a machine learning-based framework for signal detection that could complement standard statistical analyses.

Slide 3 – Research Question

The central research question I want to explore is: *Can machine learning methods uncover long-term safety signals of inclisiran that may not be detected through conventional clinical trial analyses?*

To achieve this, I propose three supporting questions.

> First, how can we harmonise structured trial data with unstructured clinical notes and national health records to create a unified dataset, which is suitable for machine learning?
>
> Second, which supervised and unsupervised learning models are most appropriate for detecting safety signals, particularly when dealing with rare or unexpected events?
>
> And third, can natural language processing of free-text notes — the kind of unstructured data routinely collected during trial visits — provide additional predictive value compared with traditional structured variables?

These questions reflect recent progress in the field of adverse drug event detection. For example, Hu *et al*. (2023) and Murphy *et al*. (2023) have shown how supervised natural language processing (or NLP) models can extract valuable safety information from clinical text. By extending these ideas to the ORION-4 dataset, we can assess whether similar techniques provide additional insights into inclisiran's long-term safety.

Slide 4 – Aims and Objectives

The overall aim of my proposal is to develop and evaluate a machine learning framework that can detect potential long-term safety signals for inclisiran.

I have set out four key objectives to achieve this aim.

First, I will undertake preprocessing of both structured datasets — such as patient demographics, laboratory values, prescriptions, and hospitalisation records — and unstructured datasets, including free-text clinical notes and adverse event reports.

Second, I will develop machine learning models, both supervised and unsupervised. Supervised models, such as random forests and neural networks, will be used to predict the risk of specific adverse events. Unsupervised clustering will be employed to look for unexpected or emerging patterns (Kim *et al*., 2022).

Third, I will benchmark these machine learning approaches against traditional statistical methods such as Cox regression survival analyses and disproportionality analysis, both of which have been widely used in clinical trial safety evaluation (Ray *et al*., 2020).

And finally, I will validate any detected signals by using external linked data from national hospitalisation and prescribing datasets, which are available for the majority of ORION-4 participants. This validation step is crucial to assess whether findings are robust beyond the immediate trial setting, and to develop methods for extended and streamlined safety assessment beyond the main trial follow-up period.

Slide 5 – Literature Review

To place this proposal in context, I have divided the literature review into two domains: cardiovascular pharmacology and machine learning in pharmacovigilance.

From the pharmacology side, statins remain the first-line treatment for hypercholesterolaemia, with PCSK9 monoclonal antibodies offered as adjuncts for very high-risk patients who do not achieve targets on statins alone (Grundy *et al*., 2019; Mach *et al*., 2020). Inclisiran represents the first RNA interference therapy licensed for lipid lowering. It was approved by the EMA in 2020 and by the FDA in 2021 (Marrs and Anderson, 2024).

Evidence from phase III trials, notably ORION-9, ORION-10, and ORION-11, demonstrated reductions in LDL cholesterol of about 50 per cent, along with a safety profile similar to placebo, aside from a higher rate of injection-site reactions (Ray *et al*., 2020). Pooled analyses and long-term extension studies, such as ORION-3 and ORION-8, confirmed sustained efficacy with no new safety concerns over approximately six years (Wright *et al*., 2023). However, the cardiovascular outcomes data clinicians are most interested in are not yet available.

On the machine learning side, a growing body of literature supports the use of both statistical learning and deep learning approaches to predict adverse drug reactions. Kim *et al*. (2022) and Murphy *et al*. (2023) have shown that models trained on electronic health records can outperform traditional approaches in certain contexts. At the same time, NLP has been used to extract adverse event information from free-text clinical notes, as demonstrated in reviews by Hu *et al*. (2023) and Murphy *et al*. (2023).

Together, this literature shows both the biological plausibility of focusing on inclisiran, given its novel mechanism, and the methodological opportunity of applying advanced machine learning and NLP to safety monitoring.

Slide 6 – Methodology

The proposed methodology will be based on integrating multiple data modalities.

For structured data, I will use trial variables such as baseline characteristics, longitudinal laboratory values, clinical observations, and serious adverse event reports. For unstructured data, I will incorporate free-text notes from trial visits, which often contain contextual details not captured in structured fields. NLP will be used to process these unstructured notes, using models such as transformer-based BERT variants that have been shown effective in biomedical text mining (Hu *et al*., 2023; Li *et al*., 2024).

The modelling strategy will involve both supervised and unsupervised approaches. Supervised classifiers such as random forests, gradient boosting, and neural networks will be developed to predict the occurrence of adverse drug events. Unsupervised clustering methods will be used to

search for novel, unexpected signals (Li *et al*., 2024; Murphy at al., 2023), in a dual approach that combines prediction of known risks with discovery of unknown patterns and can be critical for pharmacovigilance.

Slide 7 – Research Design

The research design involves several sequential steps.

First, harmonisation of structured and unstructured data, ensuring that features are aligned across sources. Feature engineering will include temporal variables, such as time since first dose, and composite measures derived from NLP-processed text.

Second, I will split the dataset into training, validation, and test sets using stratified cross-validation to ensure robust evaluation.

Third, I will benchmark machine learning models against traditional approaches, including Cox regression and disproportionality analyses (Ray *et al*., 2020), which remain the standard in clinical trials.

Finally, model evaluation will focus on metrics appropriate to imbalanced data such as rare safety events: AUROC and precision-recall curves for discrimination, calibration curves for reliability, and timeliness to assess whether machine learning can detect signals earlier than conventional methods.

Interpretability is also crucial in this context. I will apply explainable AI techniques such as SHAP values to quantify how different features contribute to predictions (Sadeghi *et al.*, 2024). This is essential to ensure clinical interpretability and build trust with stakeholders.

Slide 8 – Ethical Considerations

There are significant ethical and governance issues in using large-scale trial and national health data.

All data processing will comply with GDPR requirements. Patient identifiers will be pseudonymised, and data will be stored on secure servers with access controls.

Algorithmic bias is a recognised risk. Machine learning models may inadvertently produce spurious associations, particularly in underrepresented subgroups. Rajkomar *et al*. (2019) emphasise the importance of fairness audits to detect and mitigate these biases. I will therefore include both technical checks and expert clinical review of any signals to avoid misleading conclusions.

Ethical approval will be obtained through both the trial's governance framework and clinical research ethics boards.

Slide 9 – Timeline

The project is divided into five phases:

- Months 1 to 3 will focus on literature review, protocol finalisation, and governance approvals.
- Months 4 to 6 will involve data preprocessing, harmonisation, and development of the NLP pipeline.
- Months 7 to 9 will be dedicated to model development, including both supervised and unsupervised approaches.
- Months 10 to 12 will focus on benchmarking machine learning against standard methods and validating findings in external datasets.
- Finally, months 13 to 15 will be used for analysis consolidation, write-up, and dissemination.

This phased approach is designed to ensure feasibility within the timeframe of an MSc project.

Slide 10 – Expected Contribution

This project will deliver several important contributions.

First, it will be the first machine learning-based analysis of inclisiran safety (Marrs and Anderson, 2024; Wright *et al*., 2023). This is significant because inclisiran represents a new therapeutic class, and long-term safety data are not yet available.

Second, the project anticipates ORION-4 trial results, expected in 2026 (ClinicalTrials.gov, 2023), by providing early methodological insights and potentially detecting signals in interim or linked data.

Third, it will develop a framework for integrating structured and unstructured trial data in pharmacovigilance, which could be adapted to other trials and drugs.

Finally, the approach will be transferable to other RNA interference therapies, such as ApoC3 and ANGPTL3 inhibitors, which are currently in development (Fazoli *et al*., 2024). By establishing a robust methodology, this project can assume relevance beyond inclisiran alone.

Slide 11 – Conclusion

To conclude: Inclisiran is an effective RNA interference therapy that offers strong LDL cholesterol reduction and early indications of safety (Marrs and Anderson, 2024; Ray *et al*., 2020). However, long-term risks remain uncertain (Wright *et al*., 2023), and large-scale datasets such as ORION-4 are essential to resolve these questions (ClinicalTrials.gov, 2023).

Machine learning offers a way to complement conventional analyses by uncovering patterns that may otherwise remain hidden. Techniques such as supervised classifiers, unsupervised clustering, and natural language processing provide a rich toolkit (Kim *et al*., 2022).

The contribution of this project will therefore be both scientific and methodological: scientific, in advancing knowledge of inclisiran safety; and methodological, in demonstrating how machine learning can be used for drug safety surveillance more broadly.

Slide 12 – References

The full reference list provided here and in transcript. Thank you.

References:

ClinicalTrials.gov (2023) *A randomized trial assessing the effects of inclisiran on clinical outcomes among people with cardiovascular disease (ORION-4)*. Available at: https://clinicaltrials.gov/study/NCT03705234 (Accessed: 2 October 2025).

Fazoli, R.T. *et al.* (2024) 'RNA interference therapy in cardiology: will new targets improve therapeutic goals?', *Drugs in Context*, 13, p. 2024-3-1. Available at: 10.7573/dic.2024-3-1.

Grundy, S.M. *et al.* (2019) '2018 AHA/ACC guideline on the management of blood cholesterol', *Journal of the American College of Cardiology*, 73(24), pp. e285–e350. Available at: 10.1016/j.jacc.2018.11.003.

Hu, X. *et al.* (2023) 'Leveraging natural language processing and machine learning to identify adverse drug events in clinical text', *Drug Safety*, 46(9), pp. 815–829. Available at: 10.1007/s40264-024-01505-6.

Kim, RH. *et al.* (2022) 'Analyzing adverse drug reactions using statistical and machine learning methods: a systematic review', *Medicine*, 101(25), p. e29387. Available at: 10.1097/MD.0000000000029387.

Li, Y. *et al.* (2024) 'Artificial intelligence-powered pharmacovigilance: A review of machine and deep learning in clinical text-based adverse drug event detection for benchmark datasets', *Journal of Biomedical Informatics*, 152, p. 104621. Available at: https://doi.org/10.1016/j.jbi.2024.104621.

Mach, F. *et al.* (2020) '2019 ESC/EAS Guidelines for the management of dyslipidaemias', *European Heart Journal*, 41(1), pp. 111–188. Available at: 10.1093/eurheartj/ehz455.

Marrs, J.C. and Anderson, S.L. (2024) 'Inclisiran for the treatment of hypercholesterolaemia', *Drugs in Context*, 13, p. 2023-12-3. Available at: 10.7573/dic.2023-12-3.

Murphy, R.M. *et al.* (2023) 'Adverse drug event detection using natural language processing: A scoping review of supervised learning methods', *PloS One*, 18(1), p. e0279842. Available at: https://doi.org/10.1371/journal.pone.0279842.

Rajkomar, A. *et al.* (2019) 'Ensuring fairness in machine learning to advance health equity', *Annals of Internal Medicine*, 169(12), pp. 866–872. Available at: 10.7326/M18-1990.

Ray, K.K. *et al.* (2020) 'Two phase 3 trials of inclisiran in patients with elevated LDL cholesterol', *New England Journal of Medicine*, 382(16), pp. 1507–1519. Available at: 10.1056/NEJMoa1912387.

Sadeghi, Z. *et al.* (2024) 'A review of Explainable Artificial Intelligence in healthcare', *Computers and Electrical Engineering*, 118, p. 109370. Available at: https://doi.org/10.1016/j.compeleceng.2024.109370.

World Health Organization (2023) *Cardiovascular diseases (CVDs) fact sheet*. Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (Accessed: 2 October 2025).

Wright, R.S. *et al.* (2023) 'Safety and tolerability of inclisiran for treatment of hypercholesterolaemia: pooled analysis of 7 clinical trials', *Journal of the American College of Cardiology*, 82(24), pp. 2251–2261. Available at: 10.1016/j.jacc.2023.10.007.