

London SDE/AIC Programme: Introduction and Proposed Use-Cases

Dr. Joe Zhang (*Head of Data Science*)
Prof. James Teo (*Clinical Director AI & Data*)
Dr. Jorge Cardoso (*Chief Technology Officer*)
Jawad Chaudhry (*Programme Lead*)
Sigal Hachlili (*Director of AI, Data & Digital*)

Version 0.6 (last updated 2024 Apr 18)

Introduction

The [London AI Centre](#) (AIC) has been commissioned as part of the London Secure Data Environment (SDE) programme for its latest phase: to extend AI technologies and analytics capabilities to stakeholders and data environments across London. This document summarises the latest state of planning for the programme, as an aid to internal and external stakeholders including Integrated Care Boards (ICB) and the wider London NHS ecosystem.

What is the London SDE?

The London Secure Data Environment (SDE) is part of a national programme to enable secure and more powerful analytics for NHS, academic, and commercial users. Uniquely amongst regional peers, the London SDE does not focus on a single research platform. Rather, it places a focus on developing data infrastructure and capabilities across the region to support population health, care providers, and commissioners. This is in addition to building data environments that enable commercial research and development partnerships.

The SDE is led by **OneLondon**, as part of an overarching London Health Data Strategy, coalescing around three components (Figure 1):

- (1) **London Data Service (LDS)**: hosted in North-East London, the LDS serves as a data engineering and service layer for pan-London primary care and secondary care data. It handles data extraction and linkage, and provisions data within secure analytics environments for both research and NHS users.
- (2) **DiscoverNOW Research/Analytics Environment**: run by Imperial College Healthcare Partners in North-West London, DiscoverNOW supports governance and operation of secure research environments for academic, commercial, and NHS research and analytics.
- (3) **London AI Centre (AIC)**: a national centre of excellence for applied data science and AI, the AIC provides frontier technology for data enrichment (CogStack), federated analytics (FLIP), and deployment of machine learning tools, as well as expertise in health data and advanced analytics.

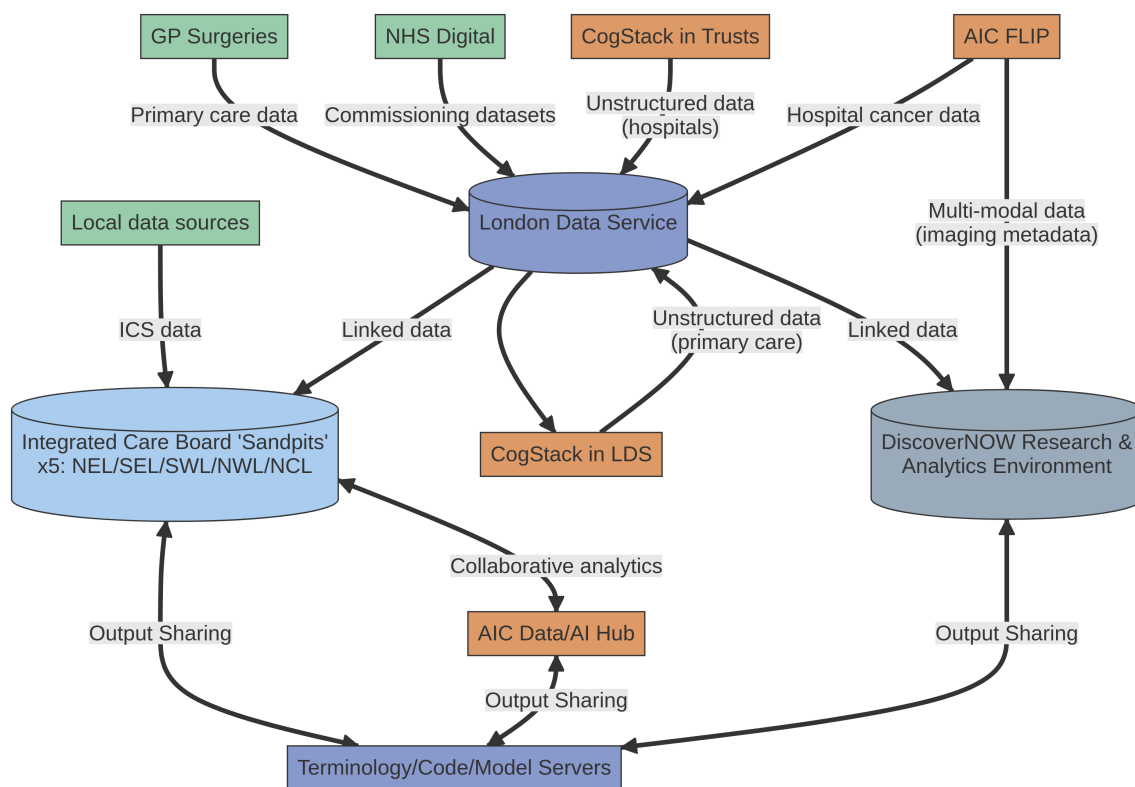


Figure 1: Summary of SDE components and data flows. Each London ICB is provisioned with its own data/analytics environment through the LDS. FLIP = Federated Learning and Interoperability Platform.

Source: [Article Notebook](#)

Technology and objectives

The contribution from the London AIC consists of technology deployment and supporting expertise, that enable a number of objectives (Figure 2) over the two year programme. This contribution includes the following:

- (1) **Federated Learning and Interoperability Platform (FLIP):** FLIP consists of (a) secure data environments within NHS hospital Trusts for multi-modal imaging data, imaging metadata, and structured health record data in the OMOP common data model; and (b) a mechanism to query data and train AI models across these secure enclaves without the need to physically transfer data. FLIP is presently installed in four major London Trusts. Integrating FLIP into the SDE will enable hospital data (such as cancer data) to be surfaced into the LDS, and enable access to multi-modal data (such as DICOM imaging and digital pathology) for research in precision healthcare.
- (2) **CogStack:** As an advanced natural language processing platform, CogStack can turn the large quantities of health information that are found in narrative text, into structured and analysable data. Currently actively used in Trusts to assist with clinical coding from notes and clinic letters, CogStack can surface secondary care and cancer pathway data, and previously unseen primary care data, into the SDE ecosystem.
- (3) **AIC Data/AI Hub:** The AIC hosts substantial health data and AI implementation expertise, that will provide practical support in data engineering, clinical informatics, data science, and machine learning (ML) development and deployment. Primary aims are to (a) help Integrated Care Boards (ICB) migrate data pipelines and analytics into common data models and terminologies within LDS environments; (b) extend these into reproducible pipelines for data science and predictive analytics deployment; and (c) work together to make ICBs self-sufficient in these capabilities. The AIC will also support the adoption and roll-out of the OMOP Common Data Model.

Source: [Article Notebook](#)

As the LDS ICB environments share a common data model, any pipelines created in collaboration with an ICB can be adapted and used for any other ICB (or deployed across multiple environments to create pan-London insights). This will also facilitate the use of shared terminologies, and validating / versioning / serving NHS-owned machine learning models across regions.

Proposed use-cases

The following use-cases are *examples* of analytics projects that can be supported within the SDE ecosystem, in collaboration between ICB/NHS analytics teams and the AIC/SDE team. Use-cases align to the London Health Data Strategy and long term condition priorities, as well as national programmes such as CORE20PLUS5, and are proposed here following early

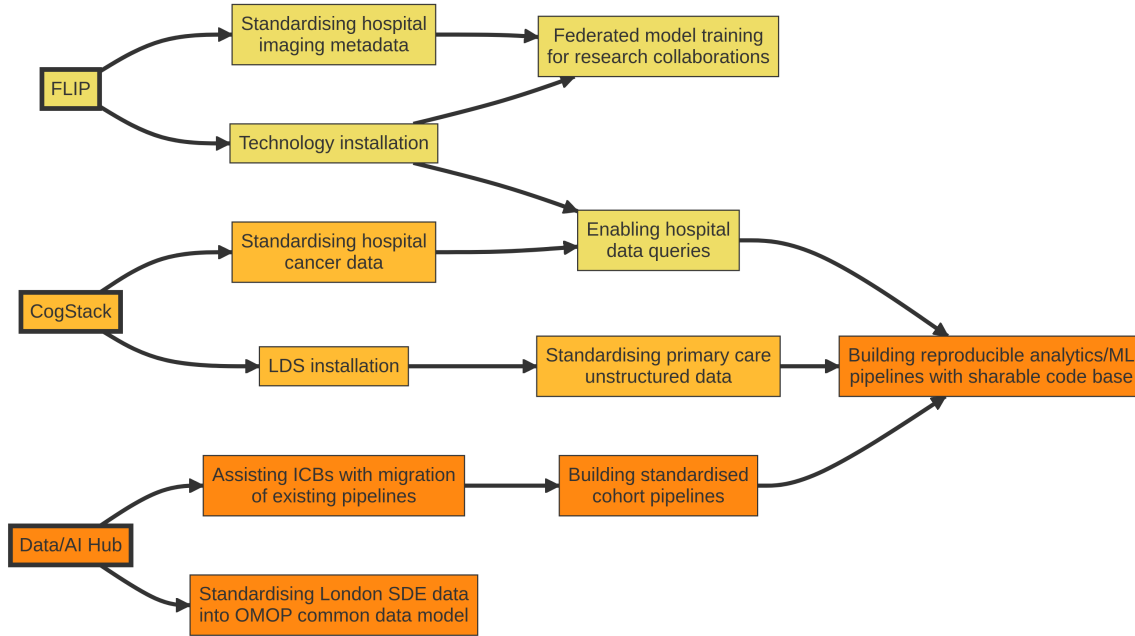


Figure 2: Summary of AIC work components and objectives. FLIP = Federated Learning and Interoperability Platform; ML = Machine Learning.

discussions with London ICBs. An overarching objective for any work is to build a foundation for reproducible analytics that can be shared across the region.

Systematic measurement of group and individual health inequality

AIM: To systematically surface multiple dimensions of health inequality across sociodemographic / geospatial groups and individual patients, and to monitor this data continuously across key long-term conditions.

BACKGROUND: Health inequality refers to measurable differences in health outcomes and determinants between individuals or groups (e.g. morbidity, co-morbidity, disease complications/death, healthcare access, disease screening, treatment delivery). Where individuals and groups experience health inequality, the principle of health *equity* emphasises the importance of reducing disparities by modifying outcome determinants that are unfairly distributed.

Health inequality is traditionally measured and visualised as a comparison of prevalence/incidence across different population groups. While helpful for broad insights, this offers limited understanding of complex individual circumstances. This type of measurement can be extended to individual patients, by using clinical domain knowledge to define ‘indicators’ of unequal disease, diagnosis, and treatment pathways. For example, in an individual with Diabetes Mellitus, indicators of inequality can include:

- (1) Diabetes surfacing at an early age;
- (2) Diagnosis in proximity to cardiovascular risk factor co-morbidities;
- (3) Diagnosis at a *late* age but with more severe disease, as measured by HbA1c or presence of end-organ complications;
- (4) Reduced health engagement/encounters/treatment compared to what is expected based on disease severity;
- (5) Shorter time to complications and mortality following diagnosis.

The precise contribution of factors to outcomes can be measured and understood in multi-variate statistical models. Overall, the presence and magnitude of indicators can be used to visualise, monitor, and explain different types of inequality, including through comparison of groups and individuals to ‘what is expected’ in a background population. The outcome is an increase in actionability, with identification of modifiable determinants of inequality (= inequality) for small groups and individuals.

APPROACH: Broadly, work would fall into three stages. The first includes defining shared terminologies, concepts, and indicators that cover long-term conditions of interest. Secondly, existing descriptions of health inequality can be migrated onto the LDS environment using shared terminologies and concepts, such that any condition can be reproducibly visualised across multiple dimensions and ‘cuts’.

Finally, this work would be extended to encompass specific inequality indicators and statistical insights, at a small group and individual level. (Figure 3) shows outcomes in an example workflow for long-term conditions (not including cancer).

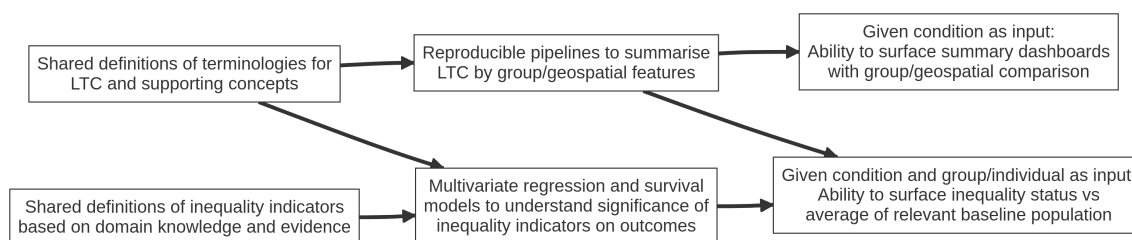


Figure 3: Outcomes in development workflow. LTC = long-term conditions.

Source: [Article Notebook](#)

The primary output of this project would be a code base that engineers cohorts from disease definitions, produces indicators for a given disease, and produces summary tables and statistics for groups and individual patients (where required). The code can be adapted by ICBs and used to support local dashboards and pathways. Code can be used for higher-level interval reporting and monitoring for the London region.

Cardiovascular disease prevention through decision intelligence

AIM: To enhance descriptive population health management with explainable predictive analytics and clinical guideline-based decision intelligence systems, across cardiovascular related co-morbidities (including hypertension, diabetes, chronic kidney disease).

BACKGROUND: The spectrum of cardiovascular long-term conditions and associated risk factors is wide, and includes hypertension, diabetes, obesity, high cholesterol, ischaemic heart disease, stroke, and chronic kidney disease, as well as dementia and heart failure. The burden of such diseases is high. Heart disease alone causes a quarter of deaths in the UK, with direct costs to the healthcare system estimated at £9 billion by the British Heart Foundation. Cardiovascular disease is seen as a **priority area for use of data** across OneLondon patient and public engagement.

There is robust aggregate understanding of cardiovascular long-term conditions in London, through prevalence reporting and Quality Outcome Framework (QOF) indicators. Existing ICB dashboards (Figure 4) routinely show how a practice or a system are performing relative to their peers. However, while indicating priority areas for action, such reporting has limitations. These include inability to surface individual patients and/or direct actions, lack of adjustment for demographics and other variables, and consideration of long-term conditions in isolation (whereas multi-morbidity changes the entire risk profile and urgency of response for individuals).

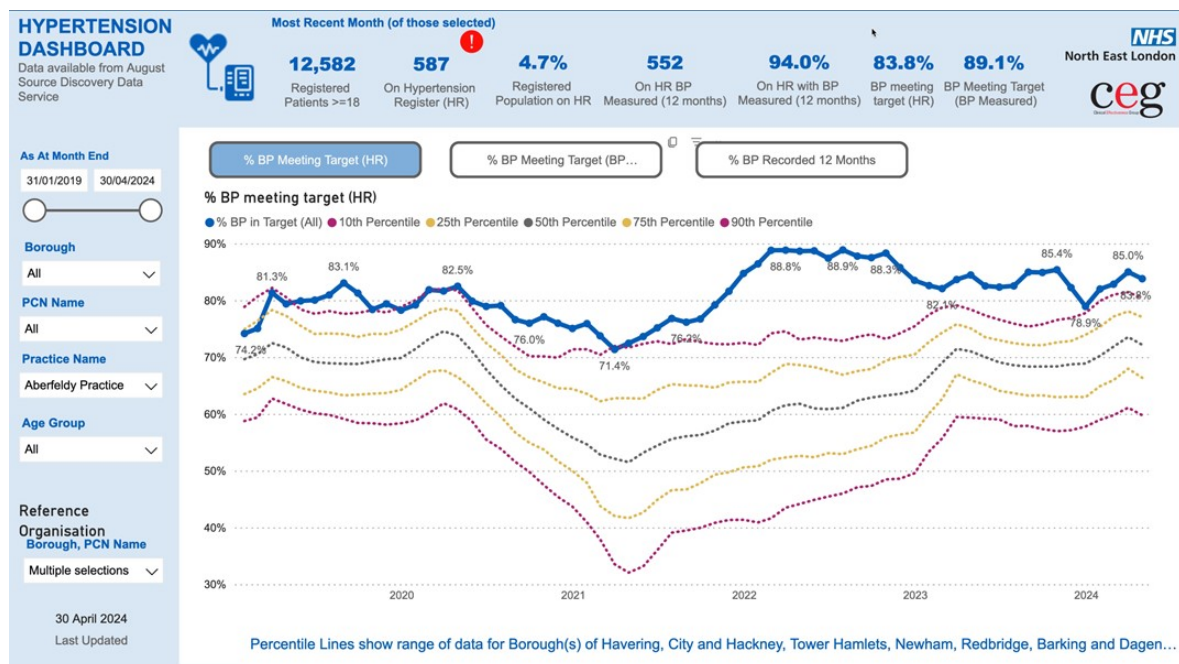


Figure 4: Existing ICB dashboard for Hypertension

Some of these limitations are being addressed by existing work in London pathfinder programmes, and in other regions such as Greater Manchester, which are moving towards electronic identification of patients who may be actioned via pre-agreed clinical pathways (Figure 5).

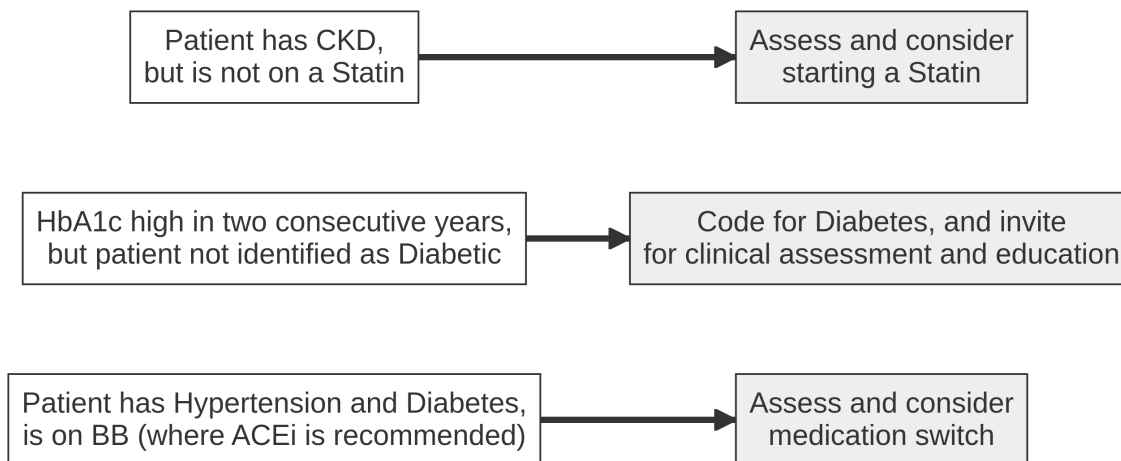


Figure 5: Examples of simple logical triggers leading to clinical actions. CKD = Chronic Kidney Disease; BB = Beta-blocker; ACEi = ACE inhibitor.

Source: [Article Notebook](#)

A previous collaboration between the AIC and North-East London ICB was able to develop precise cardiovascular risk prediction models for individuals, using explainable machine-learning algorithms and the linked patient health record. Actionable factors could also be highlighted in patients with high risk, with their relative importance explained through statistical modelling to enhance explainability (Figure 6).

APPROACH: This use-case will extend the above work, by combining multivariate statistics and machine learning for risk prediction, with robust decision systems that are grounded in evidence and clinical guidelines. Broadly, work would consist of the following components:

- (1) Development of shared terminologies, concepts, and features, used to characterise patients with any single or combination of relevant long-term condition.
- (2) Development of shared code base to ingest these definitions, and construct / describe / visualise cohorts as extension of existing dashboards. This code can be built as part of migration of existing pipelines into the LDS environment.
- (3) Computerisation of Quality Outcomes Framework targets and clinical guidelines, in conjunction with local clinical teams, to develop safe decision logic for use in the “effector” arm.

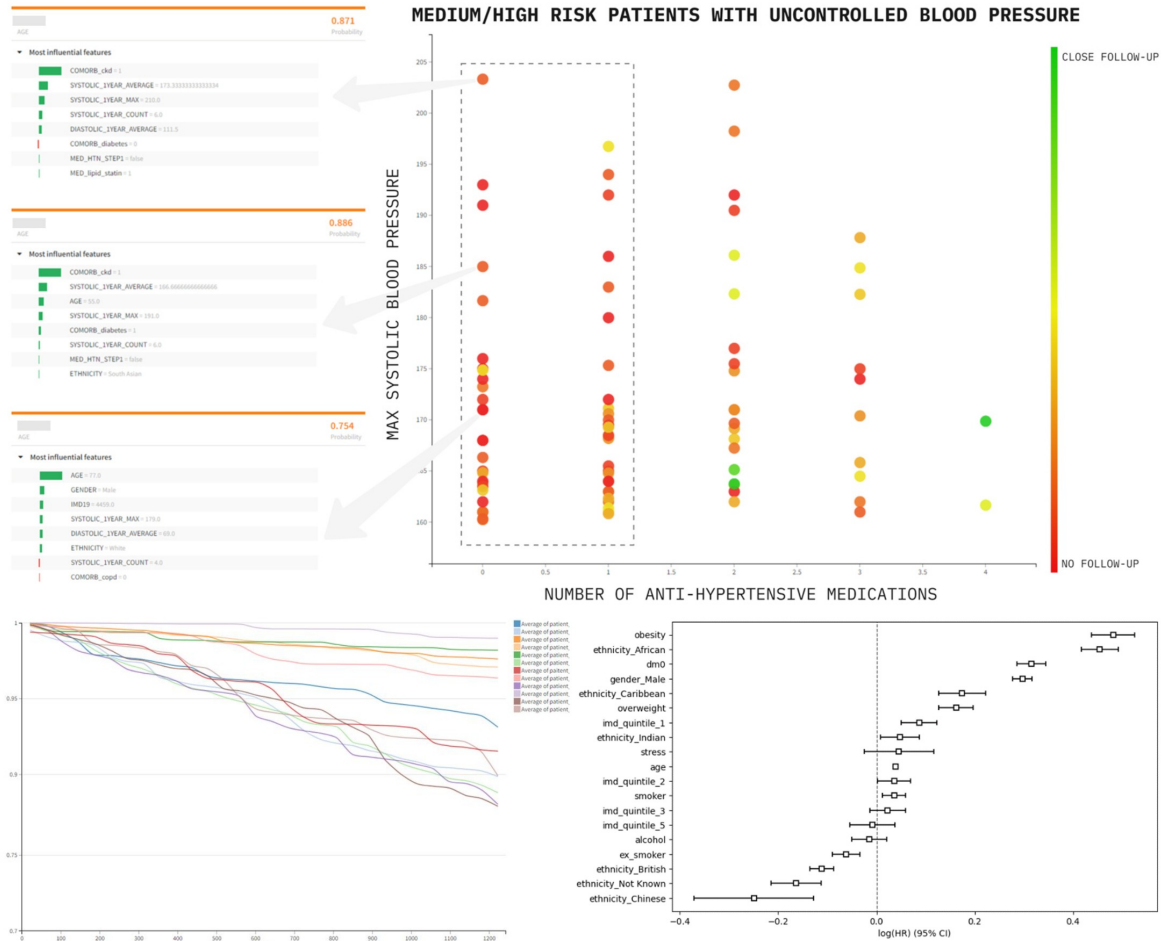


Figure 6: Actionable factors (including follow-up, treatment, blood pressure control) and association of features with adverse outcome in high risk hypertensive patients

- (4) Use of CogStack to extract additional valuable context and missing codes from unstructured text.
- (5) Development of statistical and machine learning models for predicting and understanding risk of progression across range of cardiovascular morbidity and co-morbidity.
- (6) For given patient's health record, understand actions (i.e. are there actions available, and what are they) combined with explainable risks across multiple conditions (i.e. what are the highest risks for this patient and why).

This approach aims to generate **patient-centric decision intelligence** (Figure 7), where risk and possible actions are considered for a highly detailed representation of an individual, rather than for isolated conditions, or for patients to be considered in large, aggregate groups. Any systems will need to be evaluated and monitored for safety and fairness, with a process of training and handover to continuity teams following the end of this SDE programme phase.

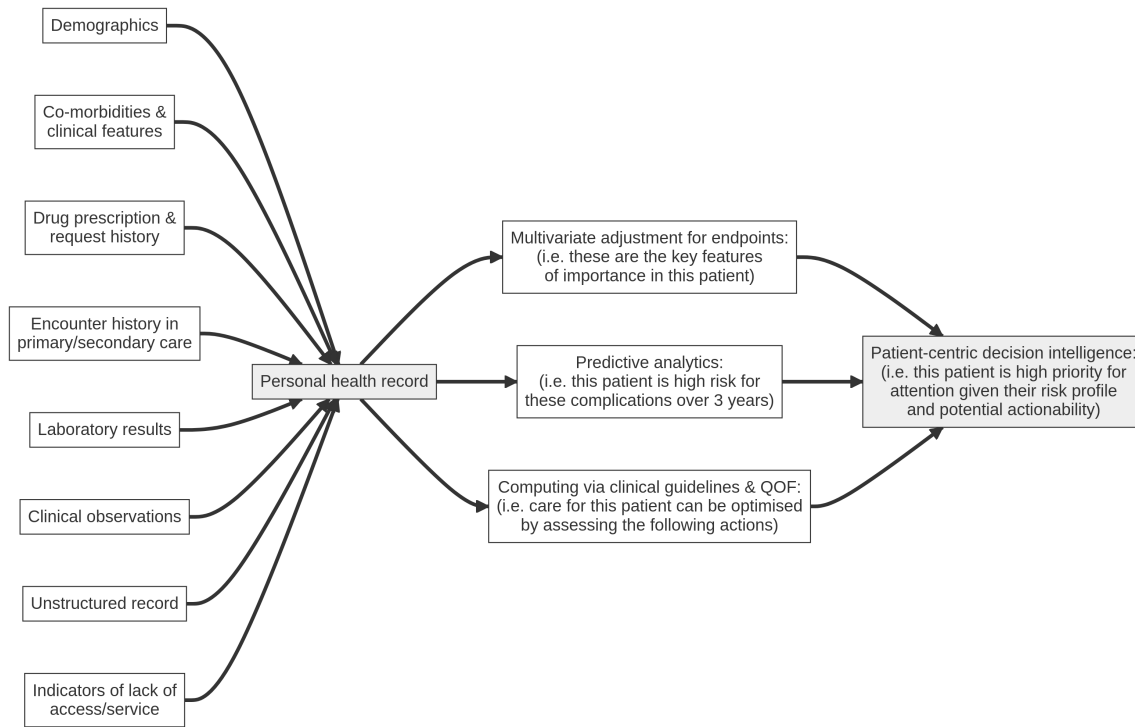


Figure 7: A linked and enriched patient healthcare record can support accurate risk prediction and prioritisation of useful actions.

Source: [Article Notebook](#)

Joining up cancer pathways

AIM:

BACKGROUND:

APPROACH:

Next steps

...