	Objective	Delivery	How will you monitor and manage success
1	1. Optimising data quality for clinical decision support (CDS)	short term (1-2 years)	Our big data analytics platform, iCARE, will provide secure and governed access to patient- level linked longitudinal data to enable research addressing data fragmentation across electronic systems and improvements in EHR data quality. We will link these data with Imperial Health Knowledge and Tissue Banks and will have contributed algorithms and knowledge into the HDRUK Phenotype Library (CALIBER/Phenoflow). We hypothesise that natural language processing (NLP) models can improve EHR data quality for appropriate/timely clinical decision-making, precision medicine and better patient outcomes. We will have tested whether NLP algorithms can derive pertinent information from clinical narratives/structured information fields held across multiple electronic systems (collaboration: SAVANA), and we will have presented outputs, at appropriate time points in the EHR, for acknowledgement/validation by health professionals, to then become a coded data point in the clinical record (collaboration: ISLA). We will have demonstrated the clinical utility of deploying NLP to capture accurate and time-specific diagnoses; comorbidities; long-term conditions; and Human Phenotype Ontology (HPO) terminology, from structured/unstructured data to enable EHR-embedded CDS. Measurable outcomes will be (1) improvements in Cerner-EHR data quality (correctness/ concordance/ defragmentation/ signal-to-noise) and (2) the 'downstream' impact on care delivery: burden on clinicians' time, rate of appropriate/timely genomic test referrals in the outpatient setting at ICHT.
2	2. Delivering an Artificial Intelligence testbed	short term (1-2 years)	To enable a secure path from development of AI-driven innovations to their trusted deployment in routine healthcare delivery, we will expand iCARE to include an AI testbed for algorithm design and testing, with user acceptance and data-protection sign-off. We will demonstrate the value of AI in detecting pattern changes in medical image data not easily amenable to human identification, by completing a project to address the need for transparency in AI image algorithms that are also fast, accurate, reproducible and applicable to both clinical trials and routine clinical practice. We will use the testbed to develop, train and test a fully automated machine learning (ML) model that can integrate complex multi-modal data and extract explainable features from brain MRI, including normalised brain volume, in patients with early multiple sclerosis (MS). We will investigate the robustness and performance of the ML model against changes of image intensity patterns and across MR scanners, and ensure it performs consistently across a diverse population to avoid algorithm bias. We will complete a retrospective cohort analysis to evaluate the association between normalised brain volume loss and healthcare resource utilisation over time in patients with MS, to predict the needs of patients and improve patient management.

3	3. Increasing real- world clinical trial capacity	short term (1-2 years)	There is a need for randomised trials to be conducted in conditions that are closer to usual clinical practice. Based on studies such as the Salford lung study, we hypothesise that our integrated healthcare records could support a system to facilitate randomised clinical trials and generate more effective real-time safety monitoring and data analytics to support the whole clinical trial lifecycle. NorthWest EHealth Limited (NWEH) are leaders in the field of electronic health records (EHRs) enabled randomised clinical trials (RCTs). They provide a feasibility and recruitment platform (FARSITE) and a clinical trials platform (ConneXon) which enable rapid feasibility and efficient recruitment, capture of EHR data for clinical trials, more effective real-time safety monitoring and data analytics to support the whole clinical trial lifecycle. They have a successful track record in developing the technology behind the Salford Lung study. Working with NWEH and Imperial College Health Partners, we will (1) expand our consultation with patients and citizens on acceptable use of their healthcare records for clinical trials held within the iCARE environment and (2) deploy clinical trial technologies for the design, recruitment and conduct of studies embedded within digital primary and secondary care clinical records.
4	4. Embedding Clinical Decision Support within the Electronic Health Record	short term (1-2 years)	Objective 1 deliverables will have demonstrated the ability to improve Cerner-EHR data quality to create a less fragmented and more usable EHR record. Building on this, we will use EHR-embedded CDS for near-real time patient risk assessment towards a reduction in patient harm events (venous thromboembolism (VTE)/falls) and risky-behaviours (smoking/alcohol). We will develop, test, and have deployed rule-based and machine learning algorithms to deliver (1) accurate and timely identification of patients with unhealthy smoking behaviours providing CDS for appropriate prescribing of nicotine replacement therapy and electronic referrals to smoking cessation community support services; (2) data-driven personalised falls risk assessment to reduce the number of falls and fall-related serious harm incidents; (3) risk-based VTE prophylaxis prescribing support to reduce in-hospital and post-discharge community-based VTE rates. Our near-real-time digital evaluation of these CDS interventions will also include evaluating clinicians' behaviours in the EHR (e.g. alert fatigue/behaviour efficiency). We will formalise our data extraction/transformation processes, metrics and measurement methodologies into a reproducible standardised digital toolkit and evaluation framework (collaboration: Oxford). Transferability will be demonstrated through deployment at five geographically diverse NHS secondary care partner sites creating federated full cycle translational analytics datasets to validate results from our analyses.

5	1. Automated interpretation of continuous cardiotocography (CTG) to improve maternal and foetal outcomes	medium term (2-3 years)	Foetal heart rate traces, cardiotocography (CTG), are challenging for health professionals to interpret visually in a continuous manner during labour. However, overinterpretation can lead to unnecessary invasive intervention (caesarean sections/instrumental deliveries); under interpretation is associated with foetal morbidity and mortality. We hypothesise that a reliable, automated AI tool can recognise and highlight episodes of non-reassuring foetal status (NRFS (i.e. foetal distress)) to support clinical decision-making during labour, towards improvements in maternal and foetal health. To address this, we will develop a predictive machine learning model for the recognition of NRFS during labour. Using our AI testbed (objective 2) and linkage of digital CTG traces (ICHT Cerner) with patient-level maternal and foetal outcomes data (delivering reusable metadata for maternity/foetal clinical analytics), we will train the algorithm to detect features of CTG tracings that signify NRFS; with learned models that are explainable to end-users for trusted application in routine care delivery. Following testing/validation of the model, algorithm outputs will be pushed into the front-end EHR for review in near-real-time by obstetricians/midwives. We will conduct a novel EHR-embedded trial to measure the impact of our AI-driven, near-real- time clinical decision support tool on the detection of NRFS and maternal/foetal morbidity and mortality.
6	2. Real World follow- up of interventions	medium term (2-3 years)	Evidence of effectiveness for interventions arising from randomised controlled trials are typically based on selected patient groups and short follow-up. Trial results are often extrapolated to inform guidelines, but may not reflect how interventions are routinely adopted in clinical practice or longer term follow-up of outcomes. Supplementing Objective 3, we hypothesise that by using linked NWL healthcare data and additional regional/national data sources (Hospital Episode Statistics) we can supplement clinical trial data to provide evidence of effectiveness in interventions where good data are lacking. By taking a population approach, we will better understand health and digital inequalities, unintended consequences and a more holistic impact on patients from healthcare interventions. We will expand our capacity to use real-world data to answer questions on intervention effectiveness and illustrate the utility of EHRs to achieve this with two exemplar projects: (1) use local EHRs to examine the relationship between non-face to face appointments and missed appointments on healthcare utilisation, digital inequalities and patient outcomes (measured through ED visits, inpatient admissions, and GP visits and actions); (2) evaluate the impact of a digital intervention for a mental health condition on adverse events and improved outcomes for a patient group in the community.

7	3. Optimising health and social care systems	medium term (2-3 years)	Health and social care systems are struggling to provide timely, safe and efficient services, particularly during a pandemic and through the post-pandemic recovery. We will develop and apply integrated machine learning and optimisation approaches for patient journeys through the healthcare system, with emphasis on improving efficient delivery of high quality and safe care and reducing inequalities.. Our approach will be data-driven and draw from quantitative logistics applications, having been recently successfully applied to hospital care prioritisation. For two exemplar patient groups (Subsets of people with mental ill-health and complex elderly patients), we will (1) use data from our Integrated Care System to complete a descriptive analysis of patient journeys through the health and social care system, characterising delays and inequalities. For our integrated health records, we will (2) formalise and document a transformation process to facilitate analysis. Taking advantage of the AI testbed delivered in objective 2, we will (3) apply our open-source framework to provide a computationally efficient approximation of the exemplar discrete optimisation problems. We will (4) publish the results of our approaches and make recommendations on how to optimise pathways based on routinely collected integrated health data, thus facilitating application to other patient groups.
8	1. Digitally enabled Clinical Decision Support for remote specialist care	long term (4-5 years)	We will implement and evaluate advanced clinical decision support (CDS) to enable EHR- embedded virtual ward round capabilities for remote specialist management of at-risk cohorts (diabetic, frail and multimorbid patients) to avoid harm (e.g. VTE/hypoglycaemic episodes) using our previously validated algorithms (short and medium-term objectives). We will extend this across the transitions of care to identify at-risk cohorts across primary, secondary and community care. These advanced CDS tools will incorporate our standardised digital toolkit and evaluation framework (objective 4) towards large-scale NHS deployment to improve patient outcomes through real time continuous evaluation demonstrating impact on quality, safety, effectiveness, and experience of care delivery. We will have completed a multiple NHS site digital intervention study to deliver translation data analytics and interpretable artificial intelligence interventions focusing on the management of falls, patient experience and patient deterioration. This will demonstrate the ability of near-real-time evaluation of interventions to measure the impact on quality, safety, effectiveness, and experience of care delivery, towards a demonstrable reduction in patient harm, across the transitions of care.

9	2. Interpretable AI applications for addressing priorities in cancer care	long term (4-5 years)	Current priorities in cancer care are driven by rising numbers of cancer patients, increased case complexity and expanding array of treatment options, and the need to improve UK cancer survival rates to be in line with the best-performing countries. We hypothesise that interpretable artificial intelligence (IAI) tools can address these priorities, by supporting (1) early diagnosis of people with suspected cancer, and (2) the identification of complex cancer cases to create adequate time for discussion of cases where it is needed and ensure the most effective use of valuable clinical and diagnostic time, in line with NHS England's guidance for cancer alliances 'Streamlining Multi-Disciplinary Team Meetings' (MDTs). Exploiting our AI testbed in iCARE (objective 2) we will use neuro-symbolic deep learning and natural language processing models to (1) develop IAI technology and tooling to support primary care physicians in making timely referrals for suspected cancer cases; (2) deliver an IAI-based decision-support tool to identify/prioritise complex cancer cases enabling MDT meeting streamlining. We will evaluate the acceptability and feasibility of IAI- driven decision support in cancer care, leading to a randomised controlled trial, in NWL, measuring impact on timeliness of cancer referrals and specialist review in two specific cancer patient cohorts.
10	3. Define approaches to Increasing the efficiency of trials embedded in routine clinical practice through interim monitoring.	long term (4-5 years)	Traditional randomised controlled trials are typically analysed after the last participant has completed their final follow up visit. We know that adaptive trials can increase efficiency by incorporating interim monitoring which can allow the trial to stop early when there is clear benefit, futility or harm. Trials embedded in routine clinical practice have the potential added advantage of being able to undertake this monitoring more frequently due to almost real- time data extraction. However, there is uncertainty around the trial design features for optimal efficiency and feasibility. We will (1) undertake an extensive simulation exercise to inform optimal trial design with respect to efficiency (smallest sample size) and develop recommendations for designing and conducting adaptive trials, including number of monitoring points/adaptive criteria when using routinely collected data. Objective 3 deliverables will have deployed clinical trial technologies for the design, recruitment and conduct of studies utilising digital healthcare data. Building on this achievement, we will (2) undertake two digital intervention trials to determine the feasibility of adaptive trial analysis and decision making, as well as validating our simulations. We will (3) generate recommendations for performing adaptive trials to improve efficiency when using routinely collected data.