# Protocol

**FULL TITLE OF THE STUDY**

Algorithm-Based Approach to CKD Identification and Its Association with Health Inequalities in Coding Practices

**SHORT STUDY TITLE / ACRONYM**

Improving Chronic Kidney Disease identification (CKD-ID)

**PROTOCOL VERSION NUMBER:** 1.1
**PROTOCOL DATE:** 10/02/2026

**RESEARCH REFERENCE NUMBERS**

| | |
|---|---|
| **IRAS Number:** | 357027 |
| **ISRCTN Number:** | To be confirmed |
| **SPONSOR:** | Lancashire Teaching Hospitals NHS Foundation Trust |

**KEY STUDY CONTACTS**

| | |
|---|---|
| Chief Investigator | Dr Wing Yin Leung, Renal Clinical Research Fellow, Renal Services, Royal Preston Hospital, Lancashire Teaching Hospitals NHS Foundation Trust, PR2 9HT, wingyin.leung@lthtr.nhs.uk |
| Collaborators | Prof Joanne Knight, Professor of Applied Data Science, Lancaster Medical School, Lancaster University, LA1 4YW, jo.knight@lancaster.ac.uk<br><br>Dr Ajay Dhaygude, Consultant Nephrologist, Renal Services, Royal Preston Hospital, Lancashire Teaching Hospitals NHS Foundation Trust, PR2 9HT, ajay.dhaygude@lthtr.nhs.uk<br><br>Dr Beng Hock So, Consultant Nephrologist, Renal Services, Royal Preston Hospital, Lancashire Teaching Hospitals NHS Foundation Trust, PR2 9HT, Beng.so@lthtr.nhs.uk<br><br>Prof Hedley Emsley, Professor of Clinical Neuroscience, Lancaster Medical School, Lancaster University, LA1 4YW, hedley.emsley@lancaster.ac.uk<br><br>Timothy Howcroft, Clinical Scientist, Lancashire Teaching Hospitals NHS Foundation Trust, PR2 9HT, timothy.howcroft@lthtr.nhs.uk |
| Sponsor & Funder | Lancashire Teaching Hospitals NHS Foundation Trust<br>Address: Royal Preston Hospital, Preston, PR2 9HT, United Kingdom |
| Lead NHS R&D contact | Paul Brown, Head of Research &Innovation, Lancashire Teaching Hospitals NHS Foundation Trust, PR2 9H, Paul.Brown@lthtr.nhs.uk |
| Statistician | Prof Joanne Knight, Professor of Applied Data Science, Lancaster Medical School, Lancaster University, LA1 4YW, jo.knight@lancaster.ac.uk |

**i. LIST of CONTENTS**



**ii. LIST OF ABBREVIATIONS**

| | |
|---|---|
| CKD | Chronic Kidney Disease |
| EHR | Electronic Health Record |
| eGFR | estimated Glomerular Filtration Rate |
| GDPR | General Data Protection Regulation |
| GP(s) | General Practitioner(s) |
| IMD | Index of Multiple Deprivation |
| ISRCTN | International Standard Randomised Controlled Trial Number |
| LTHTR | Lancashire Teaching Hospitals NHS Foundation Trust |
| L&SC / LSC | Lancashire and South Cumbria |
| LSOA | Lower-layer Super Output Area |
| NHS | National Health Service |
| NHS R&D | National Health Service Research & Development |
| NIHR | Natnioal Institute for Health and Care Research |
| OMOP | Observational Medical Outcomes Partnership |
| PCN | Primary Care Network |

| | |
|---|---|
| PCP(s) | Primary Care Provider(s) |
| PPIE | Patient and Public Involvement and Engagement |
| PSG | Patient Steering Group |
| REC | Research Ethics Committee |
| SDE | Secure Data Environment |
| uACR | urine Albumin-to-Creatinine Ratio |

**iii. STUDY SUMMARY**

<table>
<tr><td>Study Title</td><td colspan="2">Algorithm-Based Approach to CKD Identification and Its Association with Health Inequalities in Coding Practices.</td></tr>
<tr><td>Short Study Title/ Acronym</td><td colspan="2">Improving Chronic Kidney Disease identification (CKD-ID)</td></tr>
<tr><td>Study Design:</td><td colspan="2">Data-only observational study</td></tr>
<tr><td>Planned Study Period</td><td colspan="2">3 years (PhD study)</td></tr>
<tr><td></td><td>Objectives</td><td>Outcome Measures</td></tr>
<tr><td>Primary</td><td>To develop, test and refine a novel CKD detection algorithm</td><td>The diagnostic performance of the Chronic Kidney Disease (CKD) detection algorithm when applied to primary care electronic health record data</td></tr>
<tr><td>Secondary</td><td>To analyse inequalities in kidney care by comparing patients with correctly coded CKD to those with uncoded CKD</td><td>Assessment of inequalities in kidney care associated with CKD coding status in primary care</td></tr>
<tr><td>Eligibility</td><td colspan="2">Inclusion criteria<ul><li>Adults aged 18 or over</li><li>Patients registered with GP practices located within Lancashire & South Cumbria (L&SC) participating in the study.</li><li>No other specific inclusion criteria apply.</li></ul>Exclusion criteria<ul><li>Patients under the age of 18.</li><li>Patients registered with GP practices outside Lancashire & South Cumbria (L&SC).</li><li>No other exclusion criteria apply.</li></ul></td></tr>
</table>

**iv. ROLE OF STUDY SPONSOR AND FUNDER**

The study is sponsored by Lancashire Teaching Hospitals NHS Foundation Trust (LTHTR), which is also providing funding for the Chief Investigator's salary to carry out the research. LTHTR assumes overall responsibility for the initiation, management, and governance of the study. This includes oversight of regulatory compliance, information governance, and data security in line with NHS policies and applicable legislation.

The study is conducted as part of a PhD programme at Lancaster University. Lancaster University has no role as a sponsor but provides academic supervision through the Chief Investigator's supervisory team. The academic supervisors are responsible for providing methodological, statistical, and academic guidance throughout the conduct of the research.

The sponsor and Supervisory team (Dr Beng Hock So, Dr Ajay Dhaygude, Prof Hedley Emsley, Prof Joanne Knight) have contributed to the development of the study design and will provide appropriate oversight throughout the study. The Chief Investigator retains responsibility for the day-to-day conduct of the study, data analysis, interpretation of results, and preparation of manuscripts.

There is no external commercial funding for this study. Decisions relating to data analysis, reporting, and dissemination will be made independently by the research team in accordance with academic and ethical standards, and results will be disseminated regardless of outcome.

**v. ROLES AND RESPONSIBILITIES OF STUDY MANAGEMENT GROUPS**

Given the non-interventional, observational design of this study using routinely collected health records, no formal Study Steering Committee or Trial Management Group is required.

Oversight of the study will be provided by:

1) Chief Investigator: responsible for the overall conduct of the study, ensuring compliance with the protocol, ethical standards, and regulatory requirements.
2) Supervisors: providing additional oversight and guidance to the Chief Investigator.
3) Research Team meetings: regular meetings will be held to review progress, data quality, and compliance with governance and information security requirements.
4) Patient and Public Involvement and Engagement (PPIE): oversight and input will be provided through the Patient Steering Group (PSG), ensuring the study is transparent and patient centred.

**vii. KEY WORDS:**

Chronic Kidney Disease (CKD), Identification, Observational Medical Outcomes Partnership (OMOP), Coding, Health inequalities, Primary Care

## 1 BACKGROUND

Chronic Kidney Disease (CKD) is a global public health problem, as it has emerged as one of the leading causes of mortality worldwide [1]. Since 1990, we have seen significant improvement in global mortality rate for several major public health problems, however, CKD has not followed this trend and is instead predicted to become the fifth leading cause of death globally by 2040 [2]. The annual cost of kidney disease in the UK was reported to be £7.0 billion in 2023 (approximately 3.2% of National Health Service [NHS] budgets) and is expected to increase to £7.8 billion by 2033 [3]. The increasing prevalence of CKD is considered to be the biggest driver in this cost [4].

The definition and classification of CKD was first proposed in 2002 in National Kidney Foundation (NKF) Kidney Disease Outcomes Quality Initiative (KDOQI) and were subsequently adapted globally by international guideline group Kidney Disease Improving Global Outcomes (KDIGO) in 2004 [5–7]. The definition of CKD, "*kidney damage or glomerular filtration rate (GFR) <60 mL/min/1.73 m$^2$ for 3 months or more, irrespective of cause*", allowed clinicians to have a structured approach to identify and categorise kidney disease using a "traffic light" staging system [7,8].

Moderate to severe CKD affects approximately 5-6% of adults in the UK, and the prevalence is expected to rise due to the ageing population and high incidence of diseases such as hypertension and type 2 diabetes mellitus [9]. Patients with CKD have increased risk of end-stage kidney disease, cardiovascular disease, hospitalisation and death, therefore early detection of the disease is imperative to facilitate timely evaluation and management, slow progression of kidney disease, and prevent complications [10]. While patients with CKD often remain asymptomatic until later stages of the disease, CKD can be detected using blood and urine tests, measured in forms of estimated GFR (eGFR) and urine albumin to creatinine ratio (uACR).

In the UK, the majority of CKD patients are managed in primary care rather than by kidney specialists. General Practitioners (GPs) use computerised codes such as SNOMED CT, ICD-10 and retired Read codes in the Electronic Health Records (EHR) to identify these patients [11]. Those at high risk of developing CKD should receive both blood tests for eGFR and urine testing for uACR to enable accurate diagnosis and disease stratification. CKD coding on GP records triggers electronic alerts that enable systematic patient management [12], including regular blood and urine tests, blood pressure optimisation, primary prevention of cardiovascular disease and vaccinations, as recommended by the National Institute for Health and Care Excellence (NICE) guidelines and quality standards [13]. Timely implementation of these interventions slows CKD progression and avoid late presentation of end-stage kidney disease, which is associated with poorer health outcomes and higher economic burden [14,15].

Persistent discrepancies between biochemically defined CKD and its documentation in primary care registries (via GP coding) remain a significant challenge in current clinical practice, as evidenced by multiple studies [16–18]. The issue was further substantiated by the 2017 National CKD Audit (NCKDA), which offered a comprehensive assessment of CKD management in primary care across England and Wales, producing one of the largest globally representative patient cohorts with CKD [19,20]. The national audit, commissioned by the Healthcare Quality

Improvement Partnership (HQIP), aimed to improve the identification and management of CKD in primary care. One of the key issues highlighted in the report was the inaccurate assessment in CKD prevalence, where uncoded CKD ranged from 0% to 80%, suggesting a wide variation in practice performance across the 1000 audit practices. Screening for CKD in high-risk groups was inadequate, with only 54% of people with diabetes receiving relevant annual urine tests and even lower rates in other groups (such as those with hypertension), with uACR rates below 30% [19]. It was even more concerning to see that positive investigation results did not always translate into correct coding – only 70% of biochemically confirmed CKD cases were given appropriate coding on EHR, and 11% of the coded patients did not actually have CKD. Younger patients (those aged <60 years) with biochemical evidence of CKD had lower rates of coding than those aged >60 years, when this group would undoubtedly benefit from regular reviews and early specialist input given their lifetime risk of complications [19]. Linked data from primary and secondary care records also demonstrated that patients with uncoded CKD had poorer outcomes, including higher rate of unplanned hospital admission and mortality, with twice the admission rates in uncoded compared to coded patients as eGFR declines below 35 mL/min/1.73m$^2$ [20].

The audit also revealed that incorrect coding was more prevalent among ethnic minority population [19]. The UK's free healthcare system has not eliminated inequalities in kidney health, with studies consistently demonstrating higher CKD prevalence in socioeconomically deprived and ethnic minority populations [21–23]. Paradoxically, these high-risk population have lower rates of CKD coding in primary care records [24,25], resulting in delays in crucial interventions, specialist referrals and optimal disease management.

These observations of association between CKD coding and health outcomes highlight that more work needs to be done to optimise the identification and correct coding of CKD in primary care settings. The recommendations outlined in the audit were addressed not only to GPs and clinical commissioning groups (CCGs), but also to secondary care providers [20]. The report suggested that CCGs and Local Health Boards (LHBs) should consider utilising Quality Improvement (QI) tools and incentives to improve CKD care outcomes [20].

It has been eight years since NCKDA was released, yet uncoded CKD remains highly prevalent [9,15,26,27]. Traditionally, uncoded CKD patients can be identified through the health search tools built into the EHR systems using biochemical findings (eGFR and uACR). Two eGFR values (taken at least 90 days apart) are used to assess if the individual meets the diagnostic criteria for CKD. Our recent collaboration with local primary care providers in Lancashire and South Cumbria (L&SC) area suggests that incorrect coding remains a major concern. Preliminary results from the pilot study are yet to be published but show that one practice had over 100 patients (>10%) with uncoded CKD, while many coded patients did not have biochemical evidence of CKD. The Quality and Outcomes Framework (QOF) online database also demonstrated that variation in CKD prevalence between CCGs within L&SC persists, ranging from 2.46% to 5.25% [28], suggesting possible over-diagnosis of CKD in high prevalence CCGs and under-diagnosis in low prevalence CCGs. Even in practices with similar prevalence rate compared to NCKDA data, there is likely a significant number of incorrectly coded CKD cases.

Incorrectly coded patients may receive inappropriate and unnecessary intervention, whilst those with uncoded CKD may not receive appropriate targeted monitoring and management [12]. Late presentation with kidney failure increases morbidity, mortality and financial burden [29]. Many late presenters are often victims of missed opportunity. A study estimated that at least 45% of these patients could have been identified and referred for specialist care at an earlier stage [30].

There are several reasons why a patient with biochemical evidence of CKD may not be correctly coded in primary care EHR. These include limited integration of CKD case-finding tools, reliance on simplified diagnostic rules, and incomplete access to longitudinal laboratory data across care settings [31,32].

The *Fit for the Future: 10 Year Health Plan for England* (2025) sets out three major shifts in how the NHS will deliver care: moving care from hospital to community settings, harnessing digital technologies, and prioritising prevention through earlier identification of disease [[33]. These principles align closely with the aims of this project, which seeks to use digital innovation to improve early detection and accurate recording of CKD within primary care, thereby supporting proactive, community-based management. These priorities build on the earlier *NHS Long Term Plan* (2019), which emphasised prevention, early diagnosis of chronic conditions including CKD, through digital innovation and integration care systems [34]. Together, these policy frameworks highlight the need for improved, data-driven approaches to identifying CKD in real-world primary care populations. Developing and evaluating a novel algorithm to assess CKD prevalence and coding accuracy directly supports these national objectives, and the impact of such strategies warrants formal testing.

This project aims to transform CKD detection through a data-driven tool that addresses coding inaccuracies, health inequities, and economic burdens. The early stage of the study consists of three parts:

1. **Development and pilot testing** of a novel CKD detection algorithm designed to identify undiagnosed CKD cases and incorrect CKD coding.
2. **Optimisation and validation** of the algorithm to improve CKD classification accuracy by enhancing its sensitivity and specificity.
3. **Analysis of kidney care inequalities** by comparing coded and uncoded CKD groups, stratified by socioeconomic status, age, sex, ethnicity, other health conditions (comorbidities) and major complications related to CKD (sequelae) including cerebrovascular disease.

## 2 RATIONALE

Chronic Kidney Disease (CKD) is often under-recognised in primary care. Traditionally, uncoded CKD patients can be identified using search tools within or built into Electronic Health Record (EHR) systems. However, persistent discrepancies between biochemically confirmed CKD and its recording (coding) in primary care EHRs remain a major challenge. The 2017 National CKD Audit (NCKDA) showed that only 70% of patients who met the biochemical criteria for CKD were appropriately coded in their records, while 11% of those with CKD coding did not actually meet the diagnostic criteria. Linked hospital data also showed that patients without CKD coding (who

had CKD) were more likely to experience unplanned hospital admissions, acute kidney injury, and higher mortality.

There are also clear inequalities in kidney care. Despite the UK's free healthcare system, CKD is more common in people from socioeconomically deprived areas and ethnic minority backgrounds [21]. Paradoxically, these high-risk groups are less likely to have CKD correctly coded in their records [25]. This can lead to delays in diagnosis, treatment, and specialist referrals, resulting in worse outcomes. It is therefore important to evaluate whether disparities in CKD coding contribute to differences in the prevalence of major complications, including cardiovascular and cerebrovascular disease, given that CKD is a well-established risk factor for these conditions [35].

Addressing these challenges directly aligns with the *10 Year Health Plan for England: Fit for the Future (2025),* which outlines a shift in NHS care delivery from hospital to community, from analogue to digital systems, and from reactive treatment to prevention [33]. Accurate and timely identification of CKD in primary care is central to these priorities, enabling earlier intervention, maximising the benefits of digital systems, improving population health management, and supporting more equitable care through the effective use of routinely collected health data.

Together, these findings show that more effective methods are needed to improve the accuracy and completeness of CKD identification and coding in primary care.

Several new case-finding tools have been introduced in recent years, such as Ardens [36] and the Clinical Digital Resource Collaborative (CDRC) [37], which are the main tools currently used in the Lancashire and South Cumbria region. However, these tools are relatively simplistic, as they mainly rely on just two eGFR (kidney function) results when screening for CKD, without considering that most patients have many test results over time. To our knowledge, no studies have evaluated the accuracy or real-world performance of these tools.

This study will therefore develop and test a **novel CKD detection algorithm** designed to use all available kidney test results over time, with the aim of improving CKD identification and coding accuracy in primary care.

### 2.1 Assessment and management of risk

This study will use existing health records only. It will not involve any additional blood or urine tests, or the collection of information beyond what is already routinely recorded as part of standard care.

As there will be no direct patient contact for research purposes, there will be no risk of intrusion, inconvenience, or any change in the relationship between participants and their general practitioner or kidney specialists (if they are under specialist care).

The main potential risk relates to data confidentiality. To minimise this risk, strict measures will be in place throughout the study.

### 2.1.1 Risks and benefit analysis

This study aims to develop and apply a novel algorithm to existing laboratory results to improve the identification and coding of CKD. Participating general practices will provide the research team with only the NHS numbers of two cohorts of patients: (1) all patients registered with that GP practice, and (2) patients on the CKD register (those with CKD coding in their EHR). No other identifiable information will be required from GP practices. The NHS number is essential to ensure that serial laboratory results are correctly matched to the same patient and that the algorithm search is practice-specific. The research team will only access identifiable information on a strict need-to-know basis.

Demographic (including age, sex, ethnicity and socioeconomic status) and comorbidity data will be curated for part three of the study, which involves analysing kidney care inequalities by comparing coded and uncoded CKD groups. These data will be obtained from the Observational Medical Outcomes Partnership (OMOP) database. Socioeconomic status will be measured using the Index of Multiple Deprivation (IMD), derived from the individual's Lower-layer Super Output Area (LSOA) of residence. To minimise identifiability, LSOA will be used instead of postcodes.

The risk associated with handling identifiable data is mitigated through several safeguards. All data will be stored within secure, access-controlled Lancashire Teaching Hospitals NHS Foundation Trust (LTHTR) systems, using encrypted storage and transfer methods such as secure email and Microsoft shared folders linked to LTHTR accounts. Only authorised research team members who have completed mandatory Information Governance and Data Protection training will have access to the data, and only on a strict need-to-know basis. All subsequent analysis will be conducted on pseudonymised datasets. These measures ensure that identifiable information is handled securely and in accordance with the GDPR and LTHTR Information Governance policies.

The main anticipated benefit of the study is the potential for earlier and more accurate identification of CKD, with minimal burden on primary care providers and patients. Once the CKD detection algorithm is validated in real world population, the next step (not included in current protocol) will be to register the algorithm as a Class IIa medical device and implement it across Lancashire and South Cumbria. The algorithm will then be compared against current practice to assess its impact on clinical outcomes and economic implications.

Patients who are correctly coded in primary care EHRs will automatically trigger electronic alerts that enable systematic patient management, including regular blood and urine tests, blood pressure optimisation, cardiovascular risk reduction, and vaccinations, as recommended by National Institute for Health and Care Excellence (NICE) guidelines [13]. For example, prompt timely medication reviews would ensure initiation of appropriate treatments such as sodium–glucose co-transporter 2 (SGLT2) inhibitors, to help slow disease progression. Identification and coding patients who were previously missed from CKD registers may also help reduce the financial burden associated with CKD complications, including late presentation with kidney failure requiring urgent dialysis, unplanned hospital admissions, and mortality. Additionally, it may support earlier referral to secondary care services when needed.

Conversely, patients who have been incorrectly coded as having CKD could be identified and removed from the register, avoiding unnecessary interventions. This would reduce the workload for GPs and provide reassurance to patients. A cost analysis will also be conducted (not included in the current protocol) in the future to estimate savings resulting from fewer unnecessary or duplicated appointments and investigations.

### 2.1.2 Rationale for research without individual patient consent

It is not feasible to obtain individual consent due to the large scale of the population involved. For example, participating GP practices typically have approximately 10,000 registered patients each (as of November 2025) [38]. Attempting to obtain individual patient consent would be disproportionate, prohibitively costly, and could introduce significant selection bias. For these reasons, an application to the Confidentiality Advisory Group (CAG) will be part of the ethical application.

The study does not involve direct patient contact, additional samples, or the collection of data beyond what is routinely generated during standard care.

### 2.1.3 Conflict of interest

As there is no direct patient contact, risk of conflict between the researcher's clinical and research roles is minimal. Study findings will be shared with participating GP practices, stakeholders, and patient representatives to ensure transparency and inform future implementation. Ownership of the CKD detection algorithm rests with its creators, Dr Beng Hock So and Timothy Howcroft.

## 3 OBJECTIVES AND OUTCOME MEASURES

### 3.1 Objectives

This study aims to answer the key question: Can a new algorithm improve the identification of Chronic Kidney Disease (CKD) in primary care?

### 3.1.1 Primary objective

1. To develop, test and refine a novel CKD detection algorithm using longitudinal kidney test data to accurately identify CKD and optimise coding accuracy in primary care electronic health record.

### 3.1.2 Secondary objective

1. To analyse inequalities in kidney care by comparing patients who have CKD correctly recorded ("coded") in their primary care records with those who do not. The comparison will consider factors such as socioeconomic status, age, sex, ethnicity and comorbidities. We are particularly interested in whether missing CKD coding is linked to major complication related to CKD (sequelae), such as stroke or heart-related problems.

### 3.2 Outcome measures

The study outcomes focus on evaluating the diagnostic performance of a CKD detection algorithm applied to primary care electronic health record data, and on assessing inequalities in kidney care associated with CKD coding status.

The primary care outcome reflects algorithm efficacy in identifying confirmed CKD cases. Secondary outcomes explore kidney care inequality related to CKD coding and associated patient characteristics.

No adverse-effect outcomes are anticipated, as this is a data-driven study without patient contact.

#### 3.2.1 Primary outcome

The primary outcome of this study is the diagnostic performance of the Chronic Kidney Disease (CKD) detection algorithm when applied to primary care electronic health record data. This will be assessed by comparing the algorithm's classification of CKD status against a validated reference standard generated through GP–researcher co-review of longitudinal kidney test results.

Diagnostic accuracy will be quantified using standard performance metrics, including sensitivity and specificity, calculated at the point of algorithm application using all available historical kidney test data within the study period. This primary outcome reflects the efficacy of the algorithm in accurately identifying individuals with confirmed CKD and will underpin the main analysis of the study.

#### 3.2.2 Secondary outcome

Secondary outcomes focus on assessing inequalities in kidney care associated with CKD coding status in primary care. These will be evaluated by comparing patients with correctly coded CKD to those with uncoded CKD across key demographic and clinical characteristics, including socioeconomic status, age, sex, ethnicity, and relevant comorbidities, including CKD-related sequelae.

## 4 STUDY DESIGN AND METHODOLOGY

This study will look at how we can improve the way chronic kidney disease (CKD) is identified and is limited to working with data. The study does not involve direct patient contact, additional samples, or collection of data beyond what is already routinely generated during standard care.

The early stage of the study consists of three parts:

### 4.1 Part 1 of the study: Development and Pilot testing

The first part of the study involves the development and pilot testing of the CKD detection algorithm.

First, each participating GP practice is asked to provide lists of NHS numbers for two cohorts of patients registered within that individual practice: (1) all patients registered with the practice, and (2) patients included on the practice CKD register (i.e. those with CKD coding in the electronic health record [EHR]). The NHS number is essential to ensure that serial laboratory results are correctly matched to the same patient and that the algorithm search is practice-specific (i.e. restricted to individuals registered at the GP practice of interest). The research team will access identifiable information only on a strict need-to-know basis.

While CKD is defined as a sustained reduction in eGFR (estimated Glomerular Filtration Rate) of less than 60 ml/min/1.73m$^2$ for 3 months or more, most patients in real world have more than two blood tests (i.e. GFR results) over the years, and kidney function can fluctuate around the threshold of eGFR 60. To mimic how Nephrologists assess whether an individual has CKD, the algorithm will incorporate the following elements:

A. **Data Acquisition &Linkage:**
   The algorithm relies on the use of the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) which provides a standardised structure and data format to enable efficient data collection and analysis from different database systems used in primary and secondary care. The algorithm begins by extracting comprehensive laboratory data from the OMOP database, specifically targeting eGFR measurements recorded between 2015 and the time of search. The algorithm will also capture uACR (urine albumin-to-creatinine ratio) measurements and renal clinic attendance data to provide comprehensive patient profiling. The OMOP database includes laboratory results generated from both primary and secondary care settings.

B. **Preprocessing**:
   Kidney function can vary depending on a vast number of factors. For example, kidney function can drop temporarily because of an acute illness including infection, severe dehydration, and decompensated heart failure. These episodes are called Acute Kidney Injury (AKI) and may confound chronic disease assessment. To overcome this, contemporaneous eGFR values associated with AKI alerts (using specific SNOMED codes) will be removed as they reflect an acute illness rather than true assessment of the chronic disease.

C. **Temporal Analysis and Time calculation:**
   Following the data acquisition and preprocessing, the algorithm generates a chronological sequence of eGFR measurements and calculates the interval between consecutive results. It then determines how much time each patient spends with eGFR values below the threshold of 60 ml/min/1.73m$^2$.
   The interval between normal (≥60) and abnormal (<60) eGFR values will be divided to reflect the time spent in each state. When a patient transitions from normal to abnormal eGFR (<60) or vice versa, the algorithm assigns half of the intervening time to each state, providing a more accurate representation of disease chronology than a simple binary classification.

D. **Multi-timeframe CKD classification:**
The algorithm then evaluates four distinct time periods: five years, three years, eighteen months, and twelve months. For each period, it determines whether patients meet CKD criteria by calculating the percentage of time their eGFR is below 60. The primary criterion requires that patients spend more than 50% of their time below the threshold over a span of more than 90 days. Patients who meet the primary criterion are classified as having "biochemically confirmed CKD".

The algorithm incorporates several safeguards to ensure the accuracy of its analysis. For example, longer time periods (e.g., five years) are included to ensure adequate coverage of the assessment period, particularly for patients with infrequent blood tests. Additionally, it flags patients with only a single abnormal result (i.e. eGFR<60) for clinical review and repeat testing to determine whether the patient meets the criteria for CKD. The algorithm also includes a temporal requirement, ensuring that multiple low eGFR results are spaced by at least 90 days, in adherence to the CKD definition.

E. **Algorithm output:**
The algorithm generates comprehensive Excel outputs comparing its findings against GP-maintained CKD registers. Each run generates three patient lists:
   a. *Validation matches*: patients identified as having CKD by both the algorithm and the primary care record (coding).
   b. *Potential missed cases:* patients identified to have CKD according to the algorithm but not listed on the GP CKD register (coding).
   c. *Potential unnecessarily coded and further clinical review required*: patients coded for CKD on GP register but not identified as having CKD according to the algorithm.

   We will also aim to compare the performance of existing built-in EMIS search tools, Ardens, CDRC, with the novel CKD detection algorithm.

**Pilot study:**
The algorithm output will be reviewed by the participating Primary Care Providers (PCPs) and Kidney specialists to generate a validated dataset. This validated dataset will serve as the reference dataset, as it reflects mutual agreement between PCPs and Kidney specialists on CKD diagnoses based on available laboratory results. In Part 2 of the study, this reference dataset will be used to assess the accuracy of the algorithm and to support the implementation of optimisation strategies.

As additional GP practices join the study, we will extend validation by asking them and Kidney specialists to independently review patient lists without being shown how the algorithm has classified them. This approach reduces potential reviewer bias and enables the creation of a more independently verified dataset, supporting a more robust assessment of the algorithm's accuracy.

### 4.2 Part 2 of the study: Optimisation and Validation

The second part of the study includes optimisation and validation of the CKD detection algorithm to improve CKD classification accuracy, by enhancing the sensitivity and specificity of the

algorithm. We will also assess whether any improvement in the algorithm's accuracy is attributable to refinement of algorithm itself or to the availability of additional data sources (i.e. combining primary and secondary care datasets).

a. **Key optimisation strategies:**
   Several optimisation strategies will be used to optimise the algorithm's accuracy. This includes, adjusting the observation window to find the optimal timeframe for trend analysis, incorporating trajectory analysis of the latest sequential eGFR values to better identify progressive kidney function decline, and reassessing the primary criterion criteria to capture CKD cases to best mimic clinical decision making.

b. **Enhanced analytical features:**
   When uACR data is available, the system calculates the KFRE (Kidney Failure Risk Equation) scores, providing five-year risk estimates for kidney failure based on age, sex, eGFR, and albuminuria levels, as outlined by National Institute for Health and Care Excellence (NICE) guidelines.

c. **Optimisation of the algorithm:**
   The algorithm will be repeatedly tested against a validated dataset with improvement made after each cycle, until it achieves optimal balance between sensitivity and specificity, while maintaining clinical relevance for real-world decision making. To ensure the model is not overfitted we will use k-fold cross-validation.

### 4.3 Part 3 of the study: Inequalities Analyses

The third part of the study involves analysis of kidney care inequalities by comparing coded and uncoded CKD groups, including stratification by socioeconomic status, age, sex, ethnicity, and relevant clinical characteristics, including comorbidities and CKD-related sequelae. The socioeconomic status is measured using the Index of Multiple Deprivation (IMD) according to the individual's Lower-layer Super Output Area (LSOA) of residence. These data will be extracted from the OMOP database.

We will specifically examine:

a. firstly, whether uncoded CKD disproportionately affects socioeconomically deprived and ethnic minority populations, which may indicate service access barriers;
b. secondly, whether miscoded CKD (i.e. coded but without CKD) show opposite demographic pattern, indicating detection bias toward healthcare-engaged groups.
c. thirdly, how health inequalities in CKD coding influence the risk and prevalence of major sequelae, such as cerebrovascular disease, across different socioeconomic and ethnic groups.

The algorithm design aims to address and reduce these inequities by using objective biochemical criteria that are independent of a patient's level of engagement or practitioner biases. It identifies individuals who may not usually respond to routine healthcare invitation but have tests undertaken for other reasons, for example, patients presenting to Emergency Department with a fall where admission blood tests reveal impaired kidney function.

### 4.4 End of Study Definition

The end of the study will be defined as the point at which all planned data extraction, data processing, and statistical analyses specified in this protocol have been completed, and the final study dataset has been locked for reporting.

## 5 PARTICIPANT ELIGIBILITY CRITERIA

### 5.1 Inclusion criteria

- Adults aged 18 or over
- Patients registered with GP practices located within Lancashire & South Cumbria (L&SC) participating in the study.
- No other specific inclusion criteria apply.

### 5.2 Exclusion criteria

- Patients under the age of 18.
- Patients registered with GP practices outside Lancashire & South Cumbria (L&SC).
- No other exclusion criteria apply.

## 6 STUDY PROCEDURES

### 6.1 Recruitment

This research study relies on active participation from GP practices within the Lancashire and South Cumbria area. Primary Care Network (PCNs) and GP practices who have confirmed participation include:

- Chorley and South Ribble Network (PCN), which includes Buckshaw Village Health Centre (GP practice) and Adlington Medical centre (GP practice). Approximately 23,000 patients are registered with this PCN.
- Stonebridge Surgery (GP practice) – approximately 8,000 patients are registered with this GP practice.
- Garstang Medical Practice (PCN) – approximately 20,000 patients are registered with this PCN.

Further GP practices / Primary Care Network will be recruited.

#### 6.1.1 Participant identification and Screening

Participating PCNs/ GP practices who are part of the patients' existing clinical care team will have access to patient records to identify potential participants. They will provide the research team with the lists of NHS numbers for two groups of patients: all registered patients and those coded with CKD (i.e., the CKD register), to enable the CKD detection algorithm to run. No other identifiable information will be provided for the research team.

No participants will be recruited by publicity through posters, leaflets, adverts or websites.

The research team will only access identifiable information on a strict need-to-know basis. Timothy Howcroft will be the key person to run the CKD detection algorithm. He has access to all routinely collected blood and urine results from OMOP database, but will require approval for utilising these health data for research purposes. All identifiable data will be stored and shared securely using approved platforms within LTHTR, such as secure email and Microsoft shared folders linked to LTHTR accounts.

### 6.2 Consent

It is not feasible to obtain individual consent due to the large scale of the population involved. For example, GP practices have approximately 10,000 registered patients each [38]. Attempting to obtain individual patient consent would be disproportionate, prohibitively costly, and could introduce significant selection bias. For these reasons, approval from Confidentiality Advisory Group (CAG) will be sought as part of the ethical application.

The study is not considered intrusive. It does not involve direct patient contact, additional samples, or collection of data beyond what is already routinely generated during standard care.

#### 6.2.1 Rights of dissenting

Individuals who choose not to allow their data to be used for research and planning under the National Data Opt-Out (NDOO) policy will be excluded from this research study. The NDOO policy enables Primary Care Providers to identify and respect individuals' preferences automatically. Where a patient has set an opt-out choice, their data will not be included in the lists (i.e. NHS numbers of all registered patients and CKD registers) provided by GP practices to the research team.

In the unlikely event that an individual's identifiable information is mistakenly shared with the research team, a mechanism will be in place to apply the NDOO and relevant data will be immediately deleted from all systems. The incident will be reported in line with the LTHTR Information Governance Policy. The relevant Information Governance team will be notified, and corrective actions will be taken to prevent recurrence.

In addition to the NDOO, we will provide a study-specific opt-out option for individuals registered with participating GP practices/ PCNs. We will use public notices in the GP practice waiting rooms and on practice website (where available) to provide clear instructions on how to opt out, either via the GP practice receptionist or by email.

These procedures ensure that the rights of dissenting patients are fully respected and that data use remains compliant with NHS data governance requirements and the UK GDPR.

### 6.3 Description of confidential patient information

The study will use routinely collected health data from participating GP practices within Lancashire and South Cumbria and OMOP database. This includes:

- NHS number (used to link records and run the CKD detection algorithm).
- Demographic information: age, sex, ethnicity, and lower-layer super output area (LSOA) to assess socioeconomic status.

- Clinical data related to kidney function: blood test results including estimated glomerular filtration rate (eGFR), urine test results (uACR), and CKD coding in primary care electronic health records.
- Comorbidity data including relevant CKD-related sequelae.

No additional data will be collected beyond what is routinely recorded during standard care. Identifiable data (NHS numbers) will be used only as required to link pseudo-ID keys within the OMOP database, after which pseudo-IDs will be used for linkage to laboratory results, pre-processing and algorithm performance evaluation and inequalities analysis. All data will be handled in accordance with NHS and LTHTR information governance policies to ensure confidentiality and security.

### 6.4 Overview of data

1. Data extraction from GP practices:
    - Participating GP practices securely provide the NHS numbers of all registered patients and their CKD register to the research team under Section 251 support.
    - No other identifiers are provided by GP practices.
2. Pseudonymisation (NHS →pseudo-ID):
    - Upon receipt within secure LTHTR systems, all NHS numbers are immediately converted into unique pseudo-IDs under Section 251 support.
    - After this point all processing, linkage and analysis are conducted using pseudo-IDs only.
    - The file containing the NHS numbers is stored separately in an encrypted, access-restricted area within LTHTR systems.
3. Linkage &Pre-processing (pseudo-ID only):
    - Laboratory results extracted from the OMOP database are linked using pseudo-IDs to combine serial laboratory results for each patient and ensure that algorithm searches remain practice-specific.
    - Data is stored and processed within the Lancashire and South Cumbria Secure Data Environment (LSC SDE).
4. Pseudonymised Analysis Dataset Creation
    - The dataset contains pseudo-ID, serial laboratory tests, demographic variables (age, sex, ethnicity, LSOA) and comorbidity data (diagnoses).
    - This dataset is used for CKD algorithm development, performance evaluation and inequalities analysis.
    - All analyses are conducted using pseudonymised data only.
5. Output and feedback:
    - Algorithm outputs (e.g. matches, missed cases, miscoding) are initially pseudonymised.
    - When clinical verification is required, secure re-linkage (pseudo-ID →NHS number) occurs within LSC SDE under Section 251 support, to allow the original GP practice and kidney specialists to review the algorithm results.
    - No identifiable patient information leaves LSC SDE except when securely returned to the responsible GP practice.

Data security measures: All data transfers are encrypted, access is restricted to authorised research team members, and identifiable data is only used on a strict need-to-know basis.

# 7 STATISTICS AND DATA ANALYSIS

## 7.1 Sample size

This study uses routinely collected primary care electronic health record data and is not designed to test the effect of an intervention. As such, a formal statistical sample size calculation based on hypothesis testing has not been performed. Instead, the sample size is determined pragmatically by the size of the available population within participating GP practices and the requirements for robust evaluation of diagnostic algorithm performance.

The participating practices collectively include approximately 51,000 registered patients, which is expected to include a substantial number of individuals with chronic kidney disease (CKD). The sample size is considered sufficient to enable evaluation of the CKD detection algorithm's diagnostic performance, including estimation of sensitivity and specificity, and to support analyses across a representative range of age, sex, ethnicity socioeconomic status, and relevant comorbidities. It also allows meaningful comparisons between patients with correctly coded CKD and those with uncoded CKD for the assessment of inequalities in kidney care.

Recruitment of additional GP practices may be undertaken if required to ensure that the algorithm performs consistently across a broader and more diverse population. Expanding the sample would also support refinement of the algorithm to capture variation in clinical practice and patient demographics, thereby improving accuracy and generalisability. Any further recruitment will be proportionate and justified by the need to achieve statistically robust and representative findings, without unnecessary use of NHS resources.

## 7.2 Methods of analysis

Quantitative analysis:

- Primary outcome: The proportion of correctly identified CKD cases by the algorithm will be calculated for each participating GP practice or Primary Care Network. Algorithm performance will be evaluated using standard diagnostic accuracy metrics, including sensitivity and specificity. Confusion matrices will be generated to summarise classification performance. Results will be reported using descriptive statistics, including counts, percentages, means, and standard deviations, as appropriate.
- Secondary outcomes: Comparisons will be made between coded and uncoded CKD groups, as well as between correctly and incorrectly classified patients, across key demographic and clinical characteristics. Variables of interest include age, sex, ethnicity, socioeconomic status, and comorbidities. Statistical analyses will include chi-square tests for categorical variables and regression analyses to assess the association between CKD coding status and demographic or clinical factors, adjusting for potential confounders where appropriate.

# 8 DATA MANAGEMENT

## 8.1 Source of data

The study will use existing routinely collected health records from participating GP practices within Lancashire and South Cumbria. Specifically:

- Primary care electronic health record (EHR), including full list of registered patients and CKD register (coding information).
- Observational Medical Outcomes Partnership (OMOP) database for laboratory (blood and urine) test results, demographic information (date of birth [for age calculation], sex, ethnicity, LSOA/IMD) and comorbidity data including relevant CKD-related sequelae.

No additional samples or data will be collected directly from patients.

## 8.2 Data handling and record keeping

All identifiable data will be processed within secure NHS environments. Identifiable data will be stored and accessed solely via Lancashire Teaching Hospitals NHS Foundation Trust (LTHTR) systems, including secure NHS email and Microsoft shared network drives linked to LTHTR accounts. Access to research data will be restricted to authorised members of the research team only and controlled through individual LTHTR login credentials and password protection.

Data storage, processing, and analysis will also be conducted within Lancashire and South Cumbria Secure Data Environment (LSC SDE), an accredited secure platform designed for high-sensitivity NHS health data and compliant with all applicable information governance, data protection, and security standards.

All personal data will be stored electronically. No paper records will be generated or retained. Personal data will be stored only on NHS- and Trust-approved, LTHTR-managed encrypted laptops. No data will be downloaded to personal devices or USB storage media. All research team members with access to identifiable data are bound by NHS confidentiality requirements and have completed mandatory information governance and data protection training.

### 8.2.1 Pseudo-key

Immediately upon receipt within secure LTHTR systems, NHS numbers are converted into unique pseudo-IDs under Section 251 support. After this point all linkage, processing and analysis use only pseudo-IDs. Pseudo-IDs ensure that serial laboratory results are correctly matched to the same patient and that the algorithm search remains practice-specific.

The pseudo-keys will be stored in a secure, access-controlled lookup table held separately within LTHTR systems and accessible only to authorised personnel on a strict need-to-know basis.

Identifiers (such as NHS numbers) will be stored only in the initial working dataset within secure LTHTR systems and will not be included in the final research dataset used for analysis. This pseudonymisation process allows the research team to conduct analyses while retaining the

ability to re-identify individual patients via pseudo-ID keys, so that clinically relevant results can be returned to GP practices for review and validation.

All identifiable data will be stored exclusively within LTHTR's secure systems and managed in compliance with LTHTR Information Governance policies, the Data Protection Act 2018, and UK GDPR requirements. Identifiers will only be stored for maximum 12 months after completion of algorithm optimisation and validation (part 2 of the study).

### 8.3 Access to Data

The confidentiality of personal data will be maintained in accordance with the General Data Protection Regulation (GDPR), the Data Protection Act 2018, the NHS Code of Confidentiality, and Lancashire Teaching Hospitals NHS Foundation Trust (LTHTR) information governance policies.

Only authorised members of the research team, who are employed by or hold contracts with LTHTR, will have access to identifiable data. All members of the research team are bound by contractual duties of confidentiality equivalent to those of healthcare professionals and have completed mandatory NHS Information Governance and data protection training. External organisations or individuals will not have access to personal data.

Personal data will be minimised to the greatest extent possible and used only for the specific research purposes outlined in this study. Where feasible, data will be anonymised or pseudonymised before analysis. Identifiable information (e.g., NHS numbers) will only be used to link datasets.

All data will be stored and shared securely using LTHTR-approved systems, such as secure NHS email and Microsoft shared folders linked to LTHTR accounts. Access will be restricted through individual password-protected logins and audit controls. Data will not be transferred to unapproved devices, and no personal data will be stored on portable media (e.g., USB drives or personal laptops).

The research team will adhere to all relevant LTHTR and NHS data protection policies, and compliance will be overseen by the Trust's Data Protection Officer and Information Governance Department.

The research team members who will analyse the data are: Dr Wing Yin Leung, Prof Jo Knight, Timothy Howcroft, Dr Beng Hock So and Dr Ajay Dhaygude.

### 8.4 Storage and use of data after the end of the Study

After the study has ended, all research data will be securely stored within LTHTR systems in accordance with NHS Information Governance and UK GDPR requirements. Identifiable data, such as NHS numbers, will be retained only for maximum 12 months after completion of algorithm validation (part 2 of the study) to support data validation, audit, or regulatory requirements, after which they will be securely destroyed.

Pseudonymised datasets used for analysis will be retained for future research, publication, or audit purposes. All essential documents will be archived for a minimum of ten years after study completion. Archiving will be authorised by the Sponsor following submission of the End of Study report.

Access to archived data will be restricted to authorised members of the research team and governed by LTHTR Information Governance policies. All electronic data will continue to be stored on secure, password-protected LTHTR-approved systems, and no data will be transferred to unapproved devices or media.

## 9 MONITORING & AUDIT

The conduct of this study will be monitored in accordance with the policies and guidance of the research sponsor, LTHTR, and the UK Research Governance Framework.

Oversight of the study will be provided by the Chief Investigator, with support from the academic and clinical supervisors, to ensure that the research is conducted in line with the approved protocol, ethical standards, and relevant regulatory requirements. Regular internal meetings within the research team will be held to review study progress, data quality, and adherence to data management, information governance, and confidentiality procedures.

Formal on-site monitoring visits are not anticipated. Monitoring activities will primarily involve review of study documentation and datasets, conducted either remotely or internally. Periodic audits may be undertaken by LTHTR Research & Development, independently of the study investigators, to verify compliance with governance requirements, information security standards, and good research practice.

As this study involves secondary analysis of existing health records and does not involve participant recruitment, consent, or clinical interventions, monitoring will not include processes related to participant enrolment, allocation, or intervention adherence. There are no statutory pharmacovigilance requirements. Any issues identified through monitoring or audit will be documented and addressed promptly, with appropriate corrective actions implemented.

### 9.1 Scientific quality of the research

The scientific quality of this research has been reviewed within the Chief Investigator's institution and by the research team. It has also been assessed by Chief Investigator's Clinical Academic Supervisors (Dr Beng Hock So and Dr Ajay Dhaygude). The review process has involved discussions with the research team and supervisors regarding the study design, methodology, feasibility, and potential impact. Feedback was provided on the study aims, inclusion criteria, data management plans, and proposed analytical approaches, and these have been incorporated into the current protocol.

### 9.2 Statistical aspect of the research

The statistical aspects of the research have been reviewed by Professor Jo Knight and Timothy Howcroft and will continue to be responsible for reviewing the statistical aspects of the study.

They have relevant expertise in clinical data analysis and health informatics. Jo Knight is a Professor of Applied Data Science who has a significant publication and funding track record related to the analysis of data from the EHR. Tim Howcroft is a Clinical Scientist at LTHTR with specialist expertise in working with large-scale diagnostic datasets and applying data-driven analytical methods to support clinical decision-making.

Review will continue to focus on study design, data management, and analyses, including validation and assessment of accuracy of the CKD detection algorithm. Feedback provided by the reviewers has been incorporated into the current protocol to ensure appropriate statistical methodology.

## 10 ETHICAL AND REGULATORY CONSIDERATIONS

### 10.1 Research Ethics Committee (REC) review and Regulatory Compliance

Prior to the commencement of the study, a favourable ethical opinion will be sought from a Research Ethics Committee (REC) for the study protocol. No study activities will begin until REC approval has been obtained.

Any substantial amendments to the study protocol or supporting documents that require REC review will not be implemented until a favourable opinion has been granted by the REC. Where applicable, amendments will also be reviewed and approved by the relevant NHS R&D departments before they can be implemented in practice.

All correspondence with the REC, including approvals, amendment decisions, and reports, will be recorded and retained appropriately. The Chief Investigator will be responsible for submitting progress reports to the REC in accordance with the HRA guidance and for notifying the REC of the end of the study.

The lead NHS R&D contact for this research is Paul Brown, Head of Research & Innovation at LTHTR, with oversight and support provided by the North West RRDN.

### 10.2 Public and Patient Involvement and Engagement (PPIE)

Although patients are not directly recruited into this study, patient and public involvement is crucial in shaping the research project. The Patient and Public Involvement and Engagement (PPIE) element of the research will be led by Calum Reid, PPIE Coordinator at the NIHR Lancashire Clinical Research Facility. His experience will support meaningful collaboration between researchers and patient representatives.

Patient representatives from the Lay Research Group were invited to review the lay summary and the draft PPIE plan for this study. They provided valuable input on the acceptability of the research, particularly sharing their views on the proposed use of confidential patient information without consent.

To ensure that the perspectives of people living with kidney disease and the wider public are central to the study, we plan to establish a Patient Steering Group (PSG). The PSG will provide a structured

platform for underrepresented and vulnerable service users to share their experiences and priorities. Members of the PSG will play an active role in the management of the study by contributing to regular oversight meetings, shaping decisions, and advising on project direction.

The PSG will contribute to analysis particularly in interpreting findings, and support dissemination activities such as co-developing accessible summaries of results, advising on effective communication strategies, and contributing to presentations or reports aimed at patients, communities, and the wider public.

Their sustained involvement will ensure the study remains transparent, patient-centred, and aligned with the overall goal of improving early CKD detection and reducing health inequalities. Participants attending PPIE activities will be reimbursed in line with NIHR recommendations. Funding for reimbursement will be provided through Pre-application Support Funding awarded to the Chief Investigator by the NIHR Applied Research Collaboration (ARC) North West Coast. Further funding may be sought if additional or expanded PPIE activities are required during the study.

### 10.3 Insurance and Indemnity

The study is sponsored by Lancashire Teaching Hospitals NHS Foundation Trust (LTHTR). As an NHS-sponsored, non-interventional study, it is covered by the NHS Indemnity Scheme, which provides indemnity for negligent harm arising from the design, management, and conduct of the research. This includes indemnity for the sponsor and employer(s), as well as for investigators and collaborators involved in the study. No additional insurance arrangements are required.

### 10.4 Data protection and patient confidentiality

All investigators and study personnel will comply with the requirements of the Data Protection Act 2018 and UK GDPR in relation to the collection, storage, processing, and disclosure of personal data. The study will be conducted in accordance with the core principles of fair and lawful processing, minimum necessary for the purpose, accuracy, storage limitation, respect for the rights of data subjects, security and confidentiality.

Patient data will be used solely for research purposes to develop and validate a Chronic Kidney Disease (CKD) detection algorithm. The use of patient data is clearly justified, and patient and public representatives have been consulted to ensure public interest and acceptability. Only the minimum necessary data required to meet the study objectives will be accessed, including NHS numbers for data linkage, CKD register information, relevant laboratory test results, and pseudonymised demographic variables required for inequality analyses.

Data will be extracted directly from participating GP practice electronic health records and the OMOP database to ensure accuracy and completeness. Immediately upon receipt within secure LSC SDE, NHS numbers are converted into unique pseudo-IDs under Section 251 support. After this point all processing, linkage and analysis use only pseudo-IDs. The lookup table linking NHS numbers to pseudo-keys will be stored separately within secure, access-controlled LTHTR systems and will be accessible only to authorised personnel on a strict need-to-know basis.

All study data will be stored and processed on secure LSC SDE, with appropriate technical and organisational safeguards in place. Access will be limited to the minimum number of authorised members of the research team required for quality control, audit, and analysis. Any transfer of data between authorised parties will be conducted using encrypted methods and secure folders. No study data will be transferred or disclosed outside the UK or European Union.

Patients' rights will be respected throughout the study. Individuals who have exercised their National Data Opt-Out (NDOO) or Study-specific Opt-Out will not be included in the dataset, and any inadvertent inclusion will be identified and removed promptly. Identifiable data will be retained only for as long as necessary to support linkage and validation activities, after which analyses will be conducted using pseudonymised data. Study data will be retained in accordance with sponsor and NHS records management policies.

Dr Beng Hock So, Clinical Director in the Renal Department and Associate Clinical Information Officer at LTHTR, will have control of and act as the custodian for all data generated by the study.

### 10.5 Financial and other competing interests

There are no financial or other competing interests that might influence the Study design, conduct, or reporting for the Chief Investigator or current members of the research team. Individual researchers will not receive any personal payment beyond their normal salary, nor any other benefits or incentives, for involvement of this Study. Participating Primary Care Providers will not receive any payments, reimbursement of expenses, or incentives for participation.

The Chief Investigator and current collaborators hold no financial or personal interests (including shareholding, consultancy, or personal relationships) with any organisation sponsoring or supporting the research that could give rise to conflict of interest. Ownership of the CKD detection algorithm rests with its creators, Dr Beng Hock So and Timothy Howcroft.

### 10.6 Amendments

Any changes to the Study protocol, supporting documentation, or conduct of the Study will be managed in accordance with the Health Research Authority (HRA) and NHS Research & Development (R&D) guidance on amendments. Amendments may be classified as substantial or non-substantial in line with HRA definitions.

The Sponsor (LTHTR) will be responsible for determining whether a proposed amendment is substantial or non-substantial and for authorising all amendments prior to submission.

#### 10.6.1 Substantial and Non-substantial amendments

Substantial amendments (e.g. methodology, design, governance arrangements, or study objectives) will be submitted to REC, the HRA and relevant NHS R&D offices and any other regulatory bodies as required. Approval must be received before implementing any substantial amendment.

Non-substantial amendments (administrative or minor changes) will be notified to the Sponsor and communicated to participating NHS organisations where required. These amendments will be logged but do not require REC approval unless specified by HRA guidance.

### 10.6.2 Communication and Implementation

The Chief Investigator will be responsible for preparing the amendment documentation, submitting sponsor-approved substantial amendments through IRAS, ensuring all participating NHS organisations are informed of approved changes, and ensuring that no amendments are implement until all required approvals are in place.

### 10.6.3 Version Control and Documentation

A full amendment history will be stored securely and kept safely in LTHTR platform. Each protocol version will include a version number, a version date, a summary of changes, and a copy of all approved amendment forms. Only the most recent approved version of the protocol will be used for Study conduct.

## 11 DISSEMINIATION POLICY

### 11.1 Dissemination policy

The study will be registered with the ISRCTN registry as a data-only observational study once ethical approval has been obtained. Data arising from the study are owned by LTHTR. Upon completion of the study, the data will be analysed and a Final Study Report prepared which will form part of the Chief Investigator's PhD thesis.

The results of the study will be disseminated through multiple channels, including peer-reviewed scientific journals, conference presentations, publication on the study website, lay summaries designed to be accessible to patients and the public.

Lay summary of the findings will be developed in plain language and disseminated via various channels, including posters and printed/ digital summaries displayed in participating GP practices and Renal outpatient units, summaries shared through GP practice newsletters where available, presentation and updates delivered through the Patient Steering Group.

Although identifiable patient data will be used to develop and validate the CKD detection algorithm, all published results will be fully anonymised. Individual participants will not be identifiable in any publication, report, or presentation. All outputs will be reviewed by the research team prior to dissemination to ensure anonymity is preserved.

As this study uses existing health records only, there will be no direct contact with individual participants, and they will not be informed of the results individually. Participants will, however, be able to access study findings through the dissemination channels outlined above.

### 11.2 Authorship

Authorship of manuscripts and the final study report will follow the International Committee of Medical Journal Editors (ICMJE) criteria, requiring substantial contributions to study conception,

design, data analysis, interpretation, drafting, and approval of the final manuscript. Group authorship may be used where appropriate.

Professional medical writers will not be employed for this study. If external writing support is used in future publications, this will be fully acknowledged and the source of funding for such support will be transparently reported.

## 12 REFERENCES


[1] Jadoul M, Aoun M, Masimango Imani M. The major global burden of chronic kidney disease. Lancet Glob Health 2024;12:e342–3. https://doi.org/10.1016/S2214-109X(24)00050-0,.

[2] Bikbov B, Purcell C, Levey AS, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2020;395:709–33. https://doi.org/10.1016/S0140-6736(20)30045-3.

[3] NHS England. Quality and Outcomes Framework guidance for 2024/25 2024. https://www.england.nhs.uk/wp-content/uploads/2024/03/PRN01104-Quality-and-outcomes-framework-guidance-for-2024-25.pdf (accessed October 20, 2024).

[4] Kidney Research UK. Kidney disease: A UK public health emergency. The health economics of kidney disease to 2033 . 2023.

[5] National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis 2002;39:S1-266.

[6] Levey AS, Coresh J, Balk E, et al. National Kidney Foundation Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification. Ann Intern Med 2003;139. https://doi.org/10.7326/0003-4819-139-2-200307150-00013,.

[7] Levey AS, Eckardt KU, Tsukamoto Y, et al. Definition and classification of chronic kidney disease: A position statement from Kidney Disease: Improving Global Outcomes (KDIGO)z. Kidney Int 2005;67:2089–100. https://doi.org/10.1111/j.1523-1755.2005.00365.x.

[8] Stevens PE, Ahmed SB, Carrero JJ, et al. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int 2024;105:S117–314. https://doi.org/10.1016/J.KINT.2023.10.018/ASSET/210CFA23-1B32-495D-BD2B-45CB202E6E90/MAIN.ASSETS/GR10.JPG.

[9] Hirst JA, Ordóñez Mena JM, Taylor CJ, et al. Prevalence of chronic kidney disease in the community using data from OxRen: A UK population-based cohort study. British Journal of General Practice 2020;70:E285–93. https://doi.org/10.3399/BJGP20X708245,.

[10] Levey AS, Schoolwerth AC, Burrows NR, et al. Comprehensive Public Health Strategies for Preventing the Development, Progression, and Complications of CKD: Report of an Expert Panel Convened by the Centers for Disease Control and Prevention. American Journal of Kidney Diseases 2009;53:522–35. https://doi.org/10.1053/j.ajkd.2008.11.019.

[11] NHS England. Building healthcare software - clinical coding, classifications and terminology - NHS England Digital 2023. https://digital.nhs.uk/developer/guides-and-documentation/building-healthcare-software/clinical-coding-classifications-and-terminology (accessed October 4, 2024).

[12] Kim LG, Cleary F, Wheeler DC, et al. How do primary care doctors in England and Wales code and manage people with chronic kidney disease? Results from the National Chronic Kidney Disease Audit. Nephrology Dialysis Transplantation 2018;33:1373–9. https://doi.org/10.1093/NDT/GFX280.

[13] National Institute for Health and Care Excellence. Recommendations | Chronic kidney disease: assessment and management | Guidance | NICE 2021;16. https://www.nice.org.uk/guidance/ng203/chapter/Recommendations (accessed October 4, 2024).

[14] Lv J, Ehteshami P, Sarnak MJ, et al. Effects of intensive blood pressure lowering on the progression of chronic kidney disease: A systematic review and meta-analysis. CMAJ Canadian Medical Association Journal 2013;185:949–57. https://doi.org/10.1503/CMAJ.121468,.

[15] Cleary F, Kim L, Prieto-Merino D, et al. Association between practice coding of chronic kidney disease (CKD) in primary care and subsequent hospitalisations and death: a cohort analysis using national audit data. BMJ Open 2022;12:e064513. https://doi.org/10.1136/BMJOPEN-2022-064513.

[16] Jain P, Calvert M, Cockwell P, et al. The need for improved identification and accurate classification of stages 3-5 chronic kidney disease in primary care: Retrospective cohort study. PLoS One 2014;9. https://doi.org/10.1371/JOURNAL.PONE.0100831,.

[17] Chase HS, Radhakrishnan J, Shirazian S, et al. Under-documentation of chronic kidney disease in the electronic health record in outpatients. J Am Med Inform Assoc 2010;17:588. https://doi.org/10.1136/JAMIA.2009.001396.

[18] Walker N, Bankart J, Brunskill N, et al. Which factors are associated with higher rates of chronic kidney disease recording in primary care? A cross-sectional survey of GP practices. British Journal of General Practice 2011;61:203–5. https://doi.org/10.3399/BJGP11X561212,.

[19] Nitsch D, Caplin B, Hull SA, et al. National Chronic Kidney Disease Audit: National Report (Part 1) 2017. 2017.

[20] Cleary F, Kim LG, Caplin B, et al. National Chronic Kidney Disease Audit. National Report (Part 2) 2017. 2017.

[21] Caskey F, Dreyer G. KIDNEY HEALTH INEQUALITIES IN THE UK An agenda for change Kidney health inequalities in the United Kingdom: Reflecting on the past, reducing in the future. n.d.

[22] Phillips K, Hazlehurst JM, Sheppard C, et al. Inequalities in the management of diabetic kidney disease in UK primary care: A cross-sectional analysis of a large primary care database. Diabetic Medicine 2024;41:e15153. https://doi.org/10.1111/DME.15153.

[23] Hossain MP, Palmer D, Goyder E, et al. Association of Deprivation with Worse Outcomes in Chronic Kidney Disease: Findings from a Hospital-Based Cohort in the United Kingdom. Nephron Clin Pract 2012;120:c59–70. https://doi.org/10.1159/000334998.

[24] Hossain MP, Palmer D, Goyder E, et al. Social deprivation and prevalence of chronic kidney disease in the UK: workload implications for primary care. QJM: An International Journal of Medicine 2012;105:167–75. https://doi.org/10.1093/QJMED/HCR153.

[25] Molokhia M, Okoli GN, Redmond P, et al. Uncoded chronic kidney disease in primary care: A cross-sectional study of inequalities and cardiovascular disease risk management. British Journal of General Practice 2020;70. https://doi.org/10.3399/BJGP20X713105,.

[26] Stewart S, Kalra PA, Blakeman T, et al. Chronic kidney disease: detect, diagnose, disclose—a UK primary care perspective of barriers and enablers to effective kidney care. BMC Med 2024;22:1–12. https://doi.org/10.1186/S12916-024-03555-0/TABLES/3.

[27] Dolan S, Anand A, Kalra PA, et al. Uncoded chronic kidney disease prevalence in secondary care: a retrospective audit with population health implications. BMC Nephrol 2025;26:39. https://doi.org/10.1186/S12882-025-03967-X.

[28] LANCASHIRE AND SOUTH CUMBRIA ICS - QOF Database 2024. https://www.gpcontract.co.uk/browse/QE1/24#childorgs (accessed October 10, 2024).

[29] Smart NA, Titus TT. Outcomes of Early versus Late Nephrology Referral in Chronic Kidney Disease: A Systematic Review. Am J Med 2011;124:1073-1080.e2. https://doi.org/10.1016/J.AMJMED.2011.04.026.

[30] Roderick P, Jones C, Tomson C, et al. Late referral for dialysis: Improving the management of chronic renal disease. QJM: An International Journal of Medicine 2002;95:363–70. https://doi.org/10.1093/QJMED/95.6.363,.

[31] Stewart S, Kalra PA, Blakeman T, et al. Chronic kidney disease: detect, diagnose, disclose—a UK primary care perspective of barriers and enablers to effective kidney care. BMC Med 2024;22:1–12. https://doi.org/10.1186/S12916-024-03555-0/TABLES/3.

[32] Neale EP, Middleton J, Lambert K. Barriers and enablers to detection and management of chronic kidney disease in primary healthcare: a systematic review. BMC Nephrol 2020;21:83. https://doi.org/10.1186/S12882-020-01731-X.

[33] Department of Health and Social Care. Fit for the Future: The 10 Year Health Plan for England 2025. www.gov.uk/official-documents (accessed December 9, 2025).

[34] NHS. The NHS Long Term Plan 2019. www.longtermplan.nhs.uk (accessed May 30, 2025).

[35] Jankowski J, Floege J, Fliser D, et al. Cardiovascular Disease in Chronic Kidney Disease Pathophysiological Insights and Therapeutic Options. Circulation 2021;143:1157–72. https://doi.org/10.1161/CIRCULATIONAHA.120.050686;ISSUE:ISSUE:DOI.

[36] Ardens - Ardens Clinical n.d. https://ardens.org.uk/ardens-clinical/overview (accessed December 9, 2025).

[37] - Clinical Digital Resource Collaborative n.d. https://cdrc.nhs.uk/ (accessed December 9, 2025).

[38] NHS England. Patients Registered at a GP Practice - Single month view n.d. https://app.powerbi.com/view?r=eyJrIjoiNGZhOTc3ZGQtNmUwOS00M2M3LWFlZTItZjliMzNlYjExNmM5IiwidCI6IjM3YzM1NGIyLTg1YjAtNDdmNS1iMjIyLTA3YjQ4ZDc3NGVlMyJ9 (accessed December 1, 2025).

## 13 APPENDICES

**Study management / responsibilities**

| Role/ Organisation | Responsibilities |
|---|---|
| Sponsor | Overall responsibility for the initiation, management, and governance of the study. This includes oversight of regulatory compliance, information governance, and data security in line with NHS policies and applicable legislation. |
| Chief Investigator (PhD student for this study) | Overall responsibility for the day-to-day conduct of the study, protocol adherence, regulatory compliance, data analysis, interpretation of results, and dissemination including preparation of manuscripts. |
| Supervisory team | The PhD supervisory team includes academics from Lancaster University and clinicians from Lancashire Teaching Hospitals NHS Foundation Trust.<br><br>Academic Supervisors at Lancaster University provide methodological and academic oversight. They support study design, data analysis strategies, interpretation of results, and ensure alignment with PhD requirements.<br><br>Clinical Academic Supervisors (Nephrologist Consultants) at LTHTR provide clinical expertise in CKD, kidney care and coding expertise in clinical practices. They advise on clinical validity and relevance of the CKD algorithm, and support interpretation of clinical findings. |
| Primary Care Providers | Provide access to primary care EHR data via approved Secure Data Environment and validation of algorithm outputs. |
| Data Scientists | Support technical implementation, testing, and optimisation of the CKD detection algorithm; and carry out data processing, validation, and performance evaluation within the SDE. |
| PPIE coordinator | Provide patient and public perspectives to enhance relevance, acceptability, transparency, and communication of findings. |

**Data protection/confidentiality**

Approval from Confidentiality Advisory Group (CAG) will be obtained in view of using confidential patient information without obtaining individual consent. Confidentiality and security safeguards include:

- Identifiable data remain within the SDE
- Use of pseudonymised records.
- All data are stored, processed and analysed exclusively inside the SDE, with role-based access controls and audit logs.
- Any data transferred (e.g., linkage inputs) will use secure systems (NHSmail or SDE transfer protocols) in line with NHS data security standards.

**Study documentation and archiving**

Essential study documents, protocol versions, amendment history, approvals and analysis scripts will be archived by LTHFT (Sponsor) in accordance with the Sponsor's retention policy and HRA guidance.

No patient-level health records will be archived by the research team; these remain within the GP practices' clinical systems or the SDE. De-identified research datasets generated in the SDE will be retained only for the duration approved by the Sponsor, then securely deleted in accordance with SDE requirements.

Electronic project files will be version-controlled and stored securely on Sponsor-approved systems.

**Amendment History**

| Amendment No. | Protocol version no. | Date issued | Author(s) of changes | Details of changes made |
|---|---|---|---|---|
| --- | V1.0 | --- | --- | Initial version |
| 1 | V1.1 | 10 FEB 2026 | Wing Yin Leung | Updates following CAG queries |