Investigating kidney disease clinical epidemiology using routinely collected administrative data and proteomics.

Ryan Edward Aylward

A dissertation submitted jointly to the University of Bristol and University of Cape Town in accordance with the requirements for award of the degree of

Doctor of Philosophy

in the Faculty of Health Sciences

School of Population Health Sciences (University of Bristol) and Department of Medicine (University of Cape Town)

Submitted in October 2023

48, 127 words (excluding references and appendices)

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

ABSTRACT

Data collected routinely during healthcare visits and additional biospecimens collected as part of cohort study activities are invaluable to better understand kidney disease epidemiology. This thesis explores the detection and characterization of acute kidney injury (AKI), chronic kidney disease (CKD), acute-on-chronic kidney disease (A-on-CKD) and kidney disease progression using rule-based laboratory- and database-embedded algorithms and proteomic analysis.

The research includes three components. Firstly, an internal validation of the National Health Services England (NHSE) AKI detection algorithm-generated alerts received by the United Kingdom Renal Registry. Secondly, a description of the clinical epidemiology of AKI, CKD and A-on-CKD in Cape Town, South Africa, within the Provincial Health Data Centre, a health information exchange that houses administrative and clinical data about clients accessing public healthcare in the province. Lastly, proteins and biological pathways in association with CKD progression in older European adults were investigated (European Quality Study).

The implementation of the NHSE AKI detection algorithm in English laboratories was largely successful, though further investigation is required for alerts in people with CKD and alerts from a few outlying laboratories. Overall, the epidemiological findings in Cape Town shed light on the burden and characteristics of AKI, CKD and A-on-CKD in the region and challenges to research with routinely collected data in complex health systems like South Africa. In the EQUAL study, three proteins were associated with eGFR decline, potentially serving as markers of CKD progression and targets for treatment.

In conclusion, the digitome (administrative data) and proteome provided unique opportunities for detecting and understanding kidney disease, but limitations such as misclassification, missing data and inability to establish causal relationships were identified, requiring future refinements.

ACKNOWLEDGEMENTS

My interest in research was spawned in a humble hospital basement. There, Rob Freercks and Lizette van der Merwe, encouraged my endeavours to study kidney disease in the ICU and pursue a career path in research. A PhD opportunity was subtly mentioned and after a successful interview, I flew to Bristol to embark on a new adventure of methodological and personal discovery.

The Universities of Bristol and Cape Town have given me a unique opportunity and their staff have been incredibly supportive. I have received exceptional supervision from Fergus, Kate, Brian, Nicki and Yoav – to whom I am eternally grateful for their time, wisdom and open doors. Fergus and Jacqui welcomed me at Christmas and many dinners at their home, making it a little easier to be away from my own home.

I was always backed by my parents and friends (special mention to Kyla, Rod, Liani, Neil, Johan and Willem) who gave me strength and distraction when I needed it. I also made new friends along the way that shared this journey with me (Kathleen, Nina).

The Flynn family, landlords-turned-friends and boardgame-opponents, welcomed me and ensured that I was always well-fed on lemon drizzle cake. I thank them for their hospitality, kindness and friendship.

Finally, I would like to acknowledge the teams working at the UK Renal Registry (Anna Casula), Provincial Health Data Centre (Florence Phelanyane) and EQUAL Study investigators (Nick Chesnaye) who provided me with invaluable guidance.

This thesis is dedicated to all past and future imposter-sufferers.

AUTHOR DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the Universities' Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: Signed by candidate .....

Student number: 1878713 (UoB) AYLRYA001 (UCT)

DATE: 31 October 2023

CONFERENCE PROCEEDINGS

1. Consistency of alerts generated by, and implementation of, the NHS England AKI detection algorithm in English laboratories. Ryan E Aylward, Anna Casula, Nicki Tiffin, Yoav Ben-Shlomo, Brian Rayner, Kate Birnie, Fergus J Caskey. Poster presentation at UK Kidney Week, 4 – 6 June 2023, Newport, Wales.

2. Association of cardiometabolic proteins with eGFR decline in older Europeans with advanced chronic kidney disease. Ryan E Aylward, Samantha Hayward, Nicholas C Chesnaye, Claudia Torino, Maciej Szymczak, Christiane Drechsler, Friedo W Dekker, Marie Evans, Kitty J Jager, Christoph Wanner, Yoav Ben-Shlomo, Nicki Tiffin, Brian Rayner, Kate Birnie, Fergus J Caskey, for the EQUAL investigators. Moderated oral at the European Renal Association congress, 15 – 18 June 2023, Milan, Italy.

ABBREVIATIONS

AASK African American Study of Kidney diseases study

A-on-CKD Acute-on-chronic kidney disease

ADQI Acute Dialysis Quality Initiative

AGEs Advanced Glycation End Products

AKD Acute Kidney Disease

AKI Acute Kidney Injury

AKIN Acute Kidney Injury Network

ARIC Atherosclerosis Risk in Communities

ART Antiretroviral Treatment

BIS Berlin initiative Study

C1 Index Creatinine

CKD Chronic Kidney Disease

CKD-EPI Chronic Kidney Disease Epidemiology

CKD-PC Chronic Kidney Disease Prognosis Consortium

CRIC Chronic Renal Insufficiency Cohort study

CRP C-reactive protein

CVD Cardiovascular Disease

DAG Directed Acyclic Graph

DNA Deoxyribonucleic Acid

(e)GFR (estimated) Glomerular Filtration Rate

EHR Electronic Health Record

EQUAL The European Quality Study on treatment in advanced chronic kidney disease

GLOMMS Grampian Longitudinal Outcomes, Morbidity and Mortality Study (Scotland)

GLMM generalised linear mixed effects models

HAART Highly Active Antiretroviral Therapy

HIV Human Immunodeficiency Virus

HR Hazard Ratio

ICD-10 International Classification of Diseases, 10^{th} revision

ICU Intensive Care Unit

ISN International Society of Nephrology

KDIGO Kidney Diseases: Improving Global Outcomes

KDOQI Kidney Disease Outcome Quality Initiative

KF Kidney Failure

KFRE Kidney Failure Risk Equation

KFRT Kidney Failure Requiring Replacement Therapy

KRT Kidney Replacement Therapy

LIC Low Income Country

LIMS Laboratory Information Management System

LLMIC Low and Lower-Middle Income Countries

MDRD Modification of Diet in Renal Diseases

MPI Master Patient Index

NGAL Neutrophil Gelatinase Associated Lipocalin

NHLS National Health Laboratory Service (South Africa)

NHS National Health Service (United Kingdom)

NPX Normalised Protein eXpression (measure of Olink® protein on log-2 scale)

PCR Polymerase Chain Reaction

PEA Proximity Extension Assay

PHDC Provincial Health Data Centre (Western Cape, South Africa)

PMI Patient Master Index

PSA Patient Safety Alert

RAAS Renin Angiotensin Aldosterone System

RAG Red Amber Green (quality assessment score)

RIFLE Risk, Injury, Failure, Loss, End-stage

RV Reference Value

SCr Serum Creatinine

TB Tuberculosis

UKRR United Kingdom Renal Registry

THESIS STRUCTURE

I begin with a description of the function of the kidneys, how kidney function is measured and the currently accepted consensus definitions of acute kidney injury (AKI), acute kidney disease (AKD), chronic kidney disease (CKD) and kidney failure (KF). The limitations of the criteria used to define AKI will be explained, as well as the definition of kidney function recovery. Literature of previous AKI epidemiology, AKI-alert recognition and validation work, CKD epidemiology, and the biological underpinnings of CKD progression are explored. A description of various secondary data sources used in this thesis is given. **Chapter one** closes with an overview of the research aims and objectives.

Subsequently, **Chapter two** presents the methods that were used to investigate the implementation and consistency of electronic AKI alerts in English laboratories. The results and discussion of said analysis are presented in **chapter three**.

Chapter four gives the methods, results and discussion of the clinical epidemiology of AKI in the City of Cape Town, South Africa. This is followed by the methods, results and discussion of the clinical epidemiology of CKD in **chapter five** and acute-on-chronic kidney disease in the City of Cape Town, South Africa (**chapter six**).

Chapter seven presents the methods used to investigate the association of individual cardiometabolic proteins and biological pathway effects on CKD progression in older European people with advanced kidney disease referred to nephrology services. This is followed by the results of these analyses and discussion in **chapter eight**.

In the final **Chapter nine**, a summary of all findings is offered and discussed with an emphasis on challenges and future opportunities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF BOXES AND EQUATIONS

1 INTRODUCTION AND BACKGROUND

1.1 THE KIDNEYS

A kidney consists of nephrons, blood vessels, and interstitium. Each kidney – most people are born with two – contains ~1 million nephrons which are tubular in structure. Blood plasma is ultra-filtrated at the glomerular part of the nephron and concentrated in the tubular part of the nephron to form urine. The cells lining the glomerular and tubular parts are specialised in order to fulfil these functions. A ball of tightly connected blood vessels makes up the glomerulus and blood vessels interface with the nephron along its length of tubules which not only supply the kidney with the blood to be filtered, but also supply the delicate kidney environment with nutrients and oxygen, and return recovered solutes and fluid back to the systemic circulation. The interstitium comprises the cells between the nephrons and blood vessels providing structural and functional support.(1) The primary function of the kidney is to control volume, electrolyte, and acid-base homeostasis, and to excrete endogenous and exogenous toxins. Hormonal functions include secretion of erythropoietin and renin, and activation of 25-hydroxyvitamin D to 1,25 hydroxyvitamin D. (2) Figure 1. 1 shows a basic diagram of the kidney and nephron.

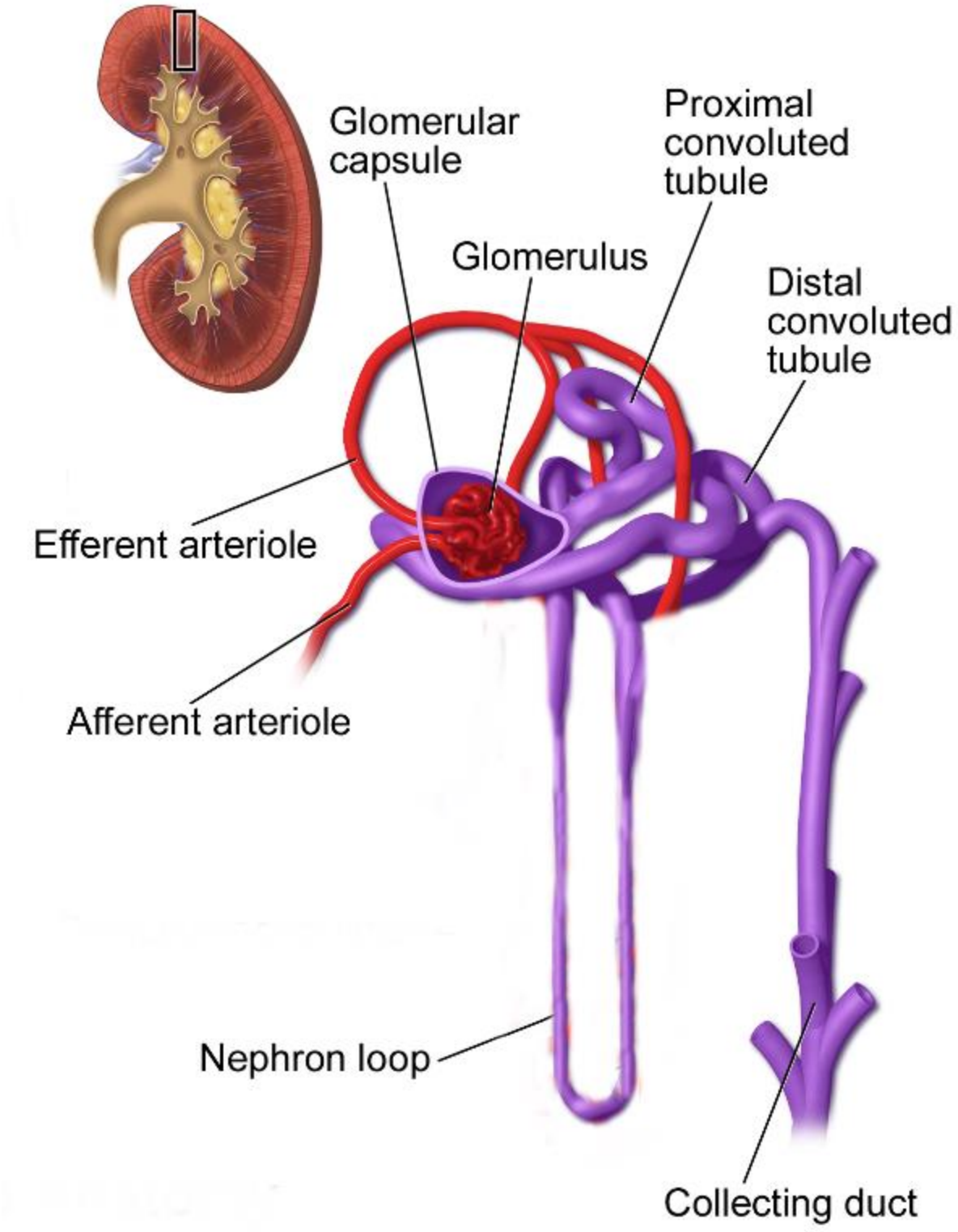


Figure 1. 1 Gross anatomy of the kidney and nephron

Reproduced under an open Creative Common licence from Visceral Organs – Advanced Anatomy 2nd. Ed. (bccampus.ca).

1.1.1 Measurement of kidney function

The filtration function of the kidney (glomerular filtration rate, GFR) cannot be measured directly and is proxied by urinary (or plasma) clearance of a filtration marker, which must be freely filtered by the glomerulus, and not reabsorbed or secreted by the tubules. It is reported in millilitres per minute, standardised to 1.73 m^2 body surface area.(2) Most of these tests are either expensive and/or difficult to perform in routine clinical practice.

Serum creatinine (SCr), a by-product of creatine from muscle catabolism, is an imperfect endogenous marker but a cheap and simple surrogate to estimate the GFR (eGFR). SCr values vary widely with extremes of body composition and age, and are affected by diet (especially red meat), exercise, supplements and drugs.(3)

Several equations to estimate GFR based on creatinine, age in years, race and sex have been developed (Modification of Diet in Renal Disease, MDRD; Chronic Kidney Disease Epidemiology Collaboration 2009 and 2021, CKD-EPI; Berlin Initiative Study, BIS).(4,5) Because the inclusion of race may promote inequalities, reduce access to kidney care and overestimate eGFR, equations without the race coefficient are now preferred in America (CKD-EPI 2021).(6) Including measurement of cystatin-C (CKD $EPI_{cr\text{-}cys}$ 2012, BIS-2) is reported to be more accurate at estimating GFR and also improves prediction of people developing kidney failure (KF) or dying because it is less influenced by factors unrelated to true GFR.(7–9) These equations are only valid in steady-state conditions due to the kinetics of creatinine production and excretion.

$$eGFR = 175 * \left(\frac{SCr}{88.4}\right)^{-1.154} * age^{-0.203} * 0.742 \textit{ [if female]} * 1.212 \textit{ [if black race]} \quad (1)$$

$$eGFR = 141 * \min\left(\frac{SCr}{k,1}\right)^{a} * \max\left(\frac{SCr}{k,1}\right)^{-1.209} * 0.993^{Age} * 1.018 \textit{ [if female]} * 1.159 \textit{ [if black race]} \quad (2)$$

$$eGFR = 142 * \min\left(\frac{SCr}{k,1}\right)^{a} * \max\left(\frac{SCr}{k,1}\right)^{-1.200} * 0.9938^{Age} * 1.012 \textit{ [if female]} \quad (3)$$

Equation 1. 1 The 4-variable MDRD (1), CKD-EPI 2009 (2), and CKD-EPI 2021 (3) equations used to estimate the glomerular filtration rate

The MDRD (1) and CKD-EPI equations (2 and 3) are the most used in clinical practice. Coefficients in the 2009 CKD-EPI equation (2): k = 0.8 and a = -0.411 for males and k =0.62 and a = -0.329 for females. Coefficients in the 2021 CKD-EPI equation (3): k = 0.7 and a = -0.241 for females, and k = 0.9 and a = -0.302 for males. The race coefficients are retained for completeness but are ignored in clinical practice. Abbreviations: **MDRD**, Modification of Diet in Renal Disease; **CKD-EPI**, Chronic Kidney Disease Epidemiology study; **eGFR**, estimated Glomerular Filtration Rate; (min)imum; (max)imum; **SCr**, serum creatinine.

1.2 SPECTRUM OF KIDNEY DISEASE

Kidney dysfunction, caused by many primary and/or secondary processes, may occur suddenly (acute kidney injury, AKI), over a moderate period (acute kidney disease, AKD) or over a long period of time leading to irreversible kidney disease (chronic kidney disease, CKD). Additionally, AKI or AKD may technically be superimposed on CKD (acute-on-chronic kidney disease). The spectrum of kidney disease may thus be seen as a continuum (Figure 1. 2).

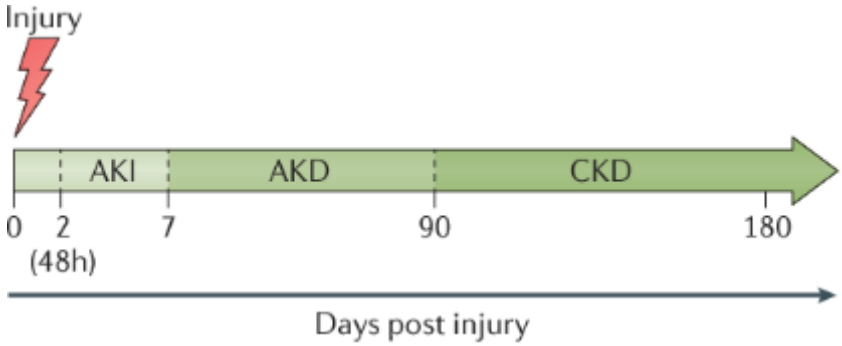


Figure 1. 2 Acute kidney injury and acute and chronic kidney disease continuum

Abbreviations: **AKI**, acute kidney injury; **AKD**, acute kidney disease; **CKD**, chronic kidney disease. Re-printed from Chawla et al under a Creative Commons license.(10)

1.2.1 Definition of acute kidney injury

AKI is defined as a sudden decrease in kidney function characterised by an increase in SCr and/or decrease in urine output with implications for health.(11) The Kidney Disease: Improving Global Outcomes (KDIGO) working group define AKI as a rise in SCr by ≥ 26.5 µmol/L (0.3mg/dL) within 48 hours or ≥ 50% within 7 days relative to baseline (see section 1.2.3) or an absolute drop in urine output <0.5ml/kg/h for a 6 hour duration. Table 1. 1 outlines the KDIGO AKI severity staging criteria which are divided into stages 1 – 3 based on SCr and urine output criteria. This consensus harmonizes the original AKI Network (AKIN) and Risk, Injury, Failure, Loss of kidney function and End stage kidney disease (RIFLE)

classifications.(12,13) The early detection of AKI is imperative if injury to the kidney is to be prevented and reversed. To this end, the International Society of Nephrology (ISN) has launched a campaign to eliminate preventable deaths attributable to AKI by 2025.(14)

Stage	SCr	Urine Output
1	1.5–1.9 times baseline within 7 days OR > 26.5 µmol/L increase in 48 hours	<0.5 ml/kg/hr for 6–12 hours
2	2.0–2.9 times baseline	<0.5 ml/kg/hr for > 12 hours
3	3.0 times baseline OR Increase in serum SCr to > 353.6 µmol/L OR Initiation of KRT OR In patients <18 years, decrease in eGFR to <35 ml/min per 1.73 m^2	<0.3 ml/kg/hr for > 24 hours OR Anuria[1] for > 12 hours

Table 1. 1 Consensus KDIGO staging classification of the severity of acute kidney injury

Redrawn from KDIGO Working Group [REF (11)]. Abbreviations: **SCr**, serum creatinine; **eGFR**, estimated Glomerular Filtration Rate; **KRT**, kidney replacement therapy; **KDIGO**, Kidney Diseases: Improving Global Outcomes.

[1] Urine output ≤ 50 ml.

1.2.2 Limitations of the current definition of acute kidney injury

Changes in SCr may reflect appropriate physiological haemodynamic responses, such as after the administration of renin angiotensin aldosterone system blockers and forced diuresis as a treatment for cardio- and hepato-renal syndromes and may actually be associated with improved outcomes.(15,16) Differences in creatinine assay or manufacturer may cause a wide variation of values within the same person. Conversely, not all injury may be detected by a rise in SCr ('subclinical AKI'), and may only be identifiable using novel biomarkers of injury and repair e.g., neutrophil gelatinase-associated lipocalin (NGAL), or kidney biopsy.(17,18) In addition, SCr rises fairly late (> 24 hours) after an insult which may lead to a delay in diagnosis of AKI.(19)

1.2.3 Baseline creatinine

The SCr level before an AKI episode is known as the baseline creatinine. Future changes in the SCr from this reference are used to define and assess severity of AKI, as described above.(11) There is no consensus on which result should be used as the reference (e.g., most recent value or nadir), whether they should be averaged, and how close in time they should be to the AKI episode.(20)

The preadmission SCr may be unavailable in as much as 50% of patients presenting with AKI.(21) When unknown, the baseline SCr may be statistically imputed in one of a number of ways: back-calculated by solving the 4-variable MDRD equation assuming an eGFR of 75ml/min/1.73m^2; using the value at the time of hospital admission, or using subsequent results after 'recovery' of the AKI episode.(20,22) These approaches are problematic,

however, as they assume baseline SCr was near-normal, there was no AKI at time of admission, or that patients are not left with AKD or CKD after an AKI episode.

1.2.4 Kidney function recovery

The definition of kidney function recovery is just as contentious and complex, with several definitions of recovery and no universal consensus.(23) KDIGO recommend a GFR ≥ 60ml/min/ 1.73m^2 be used to define complete recovery, partial recovery GFR < 60ml/min/ 1.73m^2 or dependence on kidney replacement therapy (KRT) for <90 days, and non-recovery as dependence on KRT for >90 days(11) It is argued, however, that return to normal kidney function based on SCr should not be used to signify recovery as hyperfiltration and renal reserve compensation may temporarily mask dysfunction until as much as 50% of nephrons are lost. (24)

A relative threshold for improvement, defined as return of at least 50% of the SCr level from baseline (full recovery) as suggested by the Acute Dialysis Quality Initiative (ADQI) working group, may be preferred since individuals have different baselines and may even have pre-existing CKD and therefore never reach the ≥ 60ml/min/ 1.73m^2 threshold of recovery.(12) The ADQI group also suggests that AKI that recovers within 48 hours be called *transient*, and AKI that recovers within 2 – 7 days be called *persistent*. Between 7 – 90 days, non-recovery is termed AKD. The duration of AKI before recovery occurs has been highlighted to be prognostically important.(25) Acute injury may not always recover leading to long-term kidney damage.

1.2.5 Definition of chronic kidney disease

Chronic kidney disease, defined as structural (albuminuria, findings on biopsy or ultrasound, urinary sediment) or functional (GFR) damage for longer than 90 days, can be the natural progression of AKI, especially when there are repeated insults, or the result of primary or secondary chronic diseases of the kidney, Figure 1. 2.(26) CKD is staged according to the GFR (G1 – 5) and degree of albuminuria (A1 – 3) most practically measured as urinary albumin to urinary creatinine ratio (ACR), ideally as a 24-hour urine collection or untimed early morning void, as recommended by KDIGO. This staging is based on the risk of kidney function decline and complications as highlighted by the KDIGO working group (Figure 1. 3).

The addition of albuminuria to the eGFR definition and staging of CKD is important given that albuminuria is predictive of the development of kidney failure, all-cause death and cardiovascular disease, and may herald kidney dysfunction earlier than changes in eGFR.(27) However, albuminuria testing has been shown to be infrequent in people at risk of CKD, such as those with pre-existing diabetes and hypertension.(28) Also, patients with CKD who have had albuminuria testing are less likely to be referred to nephrology services if there was significant albuminuria (> 300mg/g) versus a low eGFR (< 30ml/min/1.73m^2).(29) Investigated in primary care in the UK, albuminuria testing was only conducted in 37.0% of ambulatory patients with registered-CKD within a year after registration.(30) This may partially be explained by practices of requesting proteinuria screening instead as the test is cheaper.

Prognosis of CKD by GFR and albuminuria category

Prognosis of CKD by GFR and Albuminuria Categories: KDIGO 2012

GFR categories (ml/min/ 1.73 m²) Description and range				Persistent albuminuria categories Description and range		
				A1	A2	A3
				Normal to mildly increased	Moderately increased	Severely increased
				<30 mg/g <3 mg/mmol	30-300 mg/g 3-30 mg/mmol	>300 mg/g >30 mg/mmol
	G1	Normal or high	≥90			
	G2	Mildly decreased	60-89			
	G3a	Mildly to moderately decreased	45-59			
	G3b	Moderately to severely decreased	30-44			
	G4	Severely decreased	15-29			
	G5	Kidney failure	<15			

Green: low risk (if no other markers of kidney disease, no CKD); Yellow: moderately increased risk; Orange: high risk; Red, very high risk.

Figure 1. 3 KDIGO categories of chronic kidney disease heat map

The prognosis of CKD worsens as albuminuria increases and eGFR decreases. Reprinted from KDIGO supplements 2012 under a Creative Commons licence.(26) Abbreviations: **CKD**, chronic kidney disease; **GFR**, glomerular filtration rate; **KDIGO**, Kidney Diseases: Improving Global Outcomes; **'G'**, GFR grading; **'A',** albuminuria severity.

1.2.6 Progression of chronic kidney disease

There are several ways in which CKD progression have been defined. KDIGO defines progression as a decline in CKD stage as well as at least 25% relative reduction in GFR.(26) This ensures that an insignificant change in eGFR does not erroneously trigger a change in CKD stages across threshold boundaries e.g., an eGFR from 61 to 59 ml/min/1.73m^2. Alternatively, rapid progression is defined as a sustained decrease in eGFR > 5ml/min/1.73m^2 per year.[2] To put this into perspective, the usual rate of decline of kidney function in the general population is ~1ml/min/1.73m^2 per year starting after the 4th decade.(31)

There are several ways that the worsening in kidney function may be quantified. The change in eGFR over time, either relative (40% - 57% reduction [equivalent to doubling in SCr]), absolute slope (ml/min per 1.73m^2 per year) or a binary endpoint of reaching KF and decline in eGFR have been used.(32,33) Multiple trajectories of eGFR have been observed in people with CKD: the eGFR may decrease (CKD worsening), plateau, or increase.(32,34)

The term KF (previously *end-stage kidney disease)* is used once kidney replacement therapy (KRT; either transplantation or dialysis) is initiated or eGFR < 15 ml/min/1.73m^2 is reached and sustained for >90 days.(26,35)

KRT partially substitutes the normal functions (dialysis only replaces the filtration function) of the native kidney to sustain life. Pre-dialysis and KRT care are usually provided by specialist nephrology services that require referral.(36) Dialysis may be achieved through vascular

[2] It is noted that these two definitions, as given by the KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of CKD chapter 2.1.3, of progression are not congruous.

access using a dialysis machine (haemodialysis) at a dialysis centre or at home, or by accessing the peritoneal cavity by manual exchange of new dialysis fluid with effluent (ambulatory peritoneal dialysis) or machine (automatic peritoneal dialysis).

1.3 EPIDEMIOLOGY OF ACUTE KIDNEY INJURY

The incidence of AKI varies depending on the definition used and the setting in which it develops (in the community vs hospital vs ICU, in lower or higher income countries; in paediatric vs adult populations) and is dependent on correct human or artificial intelligence recognition.(37–39)

Susantitaphong et al conducted a world-wide meta-analysis of the burden of AKI in patients admitted to hospital which included a total of nearly 50 million individuals.(38) The included studies were diverse and used different definitions of AKI (KDIGO, other definitions and administrative coded). Most studies were conducted in high-income countries in Asia, Europe and America and consisted of hospitalised patients in critical care, cardiac surgery and hospital-acquired cases. Only 2 studies were from Africa, probably because the authors excluded studies consisting of < 500 participants and there is a general paucity of published literature from Africa. The percentage of adults with KDIGO-equivalent AKI was 21.6%, with most cases being KDIGO stage 1 (53.2%). The lower frequency of AKI when other definitions using biochemical, urine output and acute dialysis requirements, (5%) or administrative coded (2.9%) cases were used, is noteworthy.

More recently, the ISN 0by25 global snapshot included 289 centres (community-based, hospitals, or intensive care units) across 72 countries and identified 3,664 adults and 354 children with AKI from a total of 27,981 hospital inpatients screened for eligibility in a cross-sectional study.(40) Africa contributed 539 (13%) of the total number included. The median age was 60 years (IQR 43, 74; n=4,018) with a considerably younger group of patients from low-income countries (45 years old; IQR 27, 60; n=195). The average serum SCr at diagnosis was 221.9 µmol/L (IQR 146, 390) with patients from low- and lower-middle income-countries (LLMIC) averaging 288.2 µmol/L (IQR 176.8, 512.7) possibly because of delayed presentation. Using the KDIGO definition the number of cases reported in each group were: 37% in stage 1, 15% in stage 2 and 48% in stage 3. This finding – that the largest group had stage 3 disease – was likely explained by the fact that the study was a convenience sample so likely represented a highly selective group of patients of more severe referred cases compared to the study by Susantitaphong.(14,38)

The most common causes of AKI in the 0by25 study were infection/sepsis, hypotension/shock and dehydration. Nearly a quarter of all patients (n = 900, 22%) required acute dialysis. The authors conceded that this study is not completely representative as most centres were tertiary level, especially in LLMICs.

1.3.1 Acute kidney injury in Africa

There is a paucity of AKI studies from Africa. Adu and colleagues only found eight published studies conducted between 2000 and 2015 in African countries (Nigeria: 4, Sudan: 1, Burkina Faso: 1, Egypt: 1 and Ghana: 1) with a total of 703 patients.(41) Similar to the ISN Global snapshot, AKI patients tended to be relatively young (mean 41.3 years) and most developed

advanced dysfunction (56 – 80% had KDIGO stage 3 or equivalent AKI), but the risk of AKI was generally low being between 0.3 – 1.9% of admissions. Most episodes, 70 – 90%, were from the community. Nephrotoxicity (herbal and other toxins) as well as infection (sepsis, malaria) were the most common insults identified in both reviews. Obstructive uropathy from pelvic cancers (5 – 27%) and pregnancy-associated AKI (6 – 20%) were also common. Long-term CKD was not reported in the included studies, but mortality risk[3] was high, ranging between 11.5% – 43.5%.

Large, prospective observational studies of the epidemiology of AKI are therefore needed to better describe the public health burden especially in South Africa and Africa.

Considering one province in South Africa, the area where routine data are available for inclusion in this thesis, a prospective observational study of patients with AKI (n=366 between 2012 – 2013) referred to the kidney unit at Groote Schuur Hospital, Cape Town, South Africa, were also young (median age 44 years old) and 72.4% were diagnosed on admission to hospital.(42) Being a tertiary centre, most referrals were critically ill with severe AKI requiring dialysis. The reported risk[4] of 3.4%, the proportion of patients with AKI detected relative to the total number of hospital admission during the same time period, in this study is therefore likely to be an underestimate of community- and hospital-acquired AKI. Common comorbidities were HIV (20.6%), heart disease (16.1%), diabetes (17.8%) and hypertension

[3] The authors do not state at which point in time mortality was estimated e.g., 30-day mortality.
[4] The denominator population in this study were all hospital admissions registered during the same time period.

(41.5%). Sepsis and toxins were the most identified causes of AKI, consistent with other African studies.

1.3.2 Risk factors for the development of acute kidney injury

Risk factors for AKI include demographics, comorbidities, severity of illness (especially critically ill) and nature of the insult including high-risk procedures such as cardiac surgery or exposure to nephrotoxins. Since the cause of AKI is often multi-factorial, there are often many risk factors and cause additive injury to the kidney.

CKD is a strong risk for the development of AKI. The CKD Prognosis Consortium (CKD-PC), which meta-analysed 1,364,568 individual patients consisting of 22 general population and 12 CKD cohorts, highlighted the importance of lower eGFR and higher ACR at baseline on the risk of developing AKI: patients with an eGFR of 45 ml/min/1.73m^2 had a more than three-fold increased hazard compared with patients with an eGFR of 80ml/min/1.73m^2 (adjusted HR 3.35; 95% CI 2.75, 3.43) and patients with an ACR of 300mg/g had an almost three-fold increased hazard compared with patients with an ACR of less than 5mg/g (adjusted HR 2.73; 95% CI 2.18, 3.43).(43)

Advanced age and male sex have been reported as risk factors.(43,44) Comorbidities such as diabetes, hypertension, established cardiovascular disease, and HIV are also important especially because they may be potentially modifiable by effective available treatments.(43,45)

1.3.3 Nephrotoxins

Exposure to several commonly used medications (antimicrobials, non-steroidal anti-inflammatory drugs, renin angiotensin aldosterone system inhibitors, proton pump inhibitors), especially in the setting of co-occurring insults including hypovolaemia, may be risk factors for, or even the cause of, AKI.(46–48) The use of contrast media (especially intra-arterial) in the setting of hypovolaemia or pre-existing AKD or CKD is also described, but contentious.(49) Lastly, herbal preparations are important toxins in contexts like Africa and Asia where they are used culturally as traditional medicines.(50)

1.3.4 Vital outcomes

Short- and long-term survival is decreased after an AKI episode. In a systematic review, the pooled adjusted HR for mortality was 2.0 (95% CI 1.3, 3.1) in patients with AKI compared to those without.(51) The Grampian Laboratory Outcomes Morbidity and Mortality Study (GLOMMS II), consisting of over 17,000 people, observed the hazard ratio of death 1-year after an AKI episode to be the highest in hospitalized patients with baseline eGFR ≥60ml/min/ 1.73m^2 (HR 2.48; 95% CI 2.15, 2.88) and baseline eGFR 45 – 59 ml/min/ 1.73m^2 (HR 2.50; 95% CI 2.04, 3.06) compared to lower eGFR. Overall mortality was high with over 9,000 having died within a median of 9 years of follow up.(52) Mortality increases with worse AKI severity, as highlighted by Susantitaphong et al.(38)

The duration of injury also seems to be an important prognostic factor. Episodes lasting ≤ 2 days, 3 – 6 days and ≥7 days were associated with a pooled relative risk of death compared to no AKI of 1.42, 1.92 and 2.28, respectively, in a meta-analysis of over 450,000 patients.(25)

1.4 USING HEALTH INFORMATICS TO UNDERSTAND ACUTE KIDNEY INJURY EPIDEMIOLOGY

AKI epidemiology has been traditionally studied using cohort and case-control studies but may be more efficiently and comprehensively measured using large and diverse routinely electronically collected data from which retrospective or prospective cohorts and nested case-control studies can be generated. In addition, administrative data give a 'real-world' epidemiological perspective, although not without its own biases and limitations. The 2009 National Confidential Enquiry into Patient Outcomes and Death (NCEPOD) review of the care of patients with AKI in hospital showed that AKI was poorly recognised and, itself or its complications, are often preventable in the United Kingdom.(53)

AKI is notoriously under-detected and/or under-reported, especially using ICD-10 coded episodes (39,54,55) Real-time AKI electronic 'alert' intelligence systems have been developed to improve recognition and hopefully improve, AKI outcomes. They are in fact comprised of two components: the detection and the alerting process i.e., the communication of the AKI warning to clinicians which may or may not involve clinical decision support.(56) The former, the focus of this thesis, provides an important platform to study the epidemiology as datasets are large in volume, diverse (especially because not only in-hospital events are detectable), and offer greater generalisability and broader research dimensionality compared to smaller cohort studies.(56–58)

1.4.1 Acute kidney injury recognition

Algorithms use logic rules based on changes in SCr to detect AKI. Campbell et al found that a laboratory algorithm based on KDIGO criteria detected AKI in 12.4% (46,101/370,969) of admissions in an Australian centre, although AKI status was unknown in a further 22.1% due to an unknown baseline SCr.(39) Most episodes were classified as stage 1 (72.1%). Only 15.9%, 38.5% and 46.8% of laboratory identified KDIGO 1, 2 and 3 episodes, respectively, were recorded in administrative coded data. Code-identified AKI therefore tends to capture episodes that are most severe, underestimating 'milder', but still prognostically important injury.(54)

Adult patients with AKI identified using the National Health Service (NHS) England mandated algorithm have been studied at Tygerberg Hospital, Cape Town, South Africa.(59) Out of 18,781 admissions over a 6-month period, community- and hospital-acquired AKI was observed in 14.7% of patients admitted. This study had significant limitations as it was restricted to a single tertiary hospital and only a subset (n=80) of the 1,165 patients with hospital-acquired AKI were clinically characterized using medical record review.

Other algorithms have also been evaluated (Table 1. 2). One of the advantages of a rule-based algorithm is the ability to delineate AKI episodes (multiple alerts fired within a particular time frame) so that those with a single versus multiple episodes may be studied.(60) Of caution, the reference creatinine will 'reset' because the algorithm may use a new higher baseline SCr as time progresses, therefore underestimating future AKI alerts.

Although algorithms may be based on KGIDO or other criteria, there are usually pragmatic decisions that are taken to select which SCr result should be used as the baseline if there are no results within the 7 days preceding the AKI episode (something not specifically addressed by KDIGO). Laboratory systems may be limited by what may be possible within their logic or linkage to laboratory data. For example, clinical information, such as urine output or need for KRT, which are used to stage AKI, may not be available to strengthen the diagnosis of AKI. (60) This results in slight differences in implementation of any algorithm.

A major limitation of most algorithms is that they do not include urine output criteria, data which are often not collected or captured outside of the ICU, that improve the sensitivity of AKI detection.(61,62) The initiation of acute dialysis is included in KDIGO guidelines to assess stage 3 AKI.(11) This information may not always be available in routinely collected records or is not linked to biochemistry data, and is notably not featured in the NHS England algorithm.(63) Others have developed rules which are not consensus based, such as changes in eGFR instead of creatinine.(64)

	Algorithm			
	Surrey	**Nottingham**	**Pittsburgh**	**NHS England**
Author (Ref)	Tirunagari (64)	Porter (65)	Sakhuja (66)	Selby (63)
Country	England	England	United States	England
Year published	2017	2014	2021	2015
Setting	Primary care	Hospital	Hospital	In and out-patient
Study type	Retrospective	Prospective	Prospective	Retrospective
Number of participants	488	15,550	337,380	49,718/ 4,464
Criteria	MDRD eGFR 1.5x increase	RIFLE, AKIN*	50 % rise in previous year or 0.3mg/dL rise above last value within 52 hour and 0.3mg/dL rise above baseline	KDIGO
Baseline kidney function	Not specified	Nadir 7 – 365 days, eGFR 75 back-calculation when unknown	Median 0 – 365 days before admission	Lowest within past 7 days, median within 8 – 365 days
Exclusions	Not specified	Chronic dialysis, age < 16 years	KF, acute dialysis	No algorithm exclusions§
Gold standard	Nephrologist adjudication	Not applicable	Nephrologist adjudication	ICD-10 coded AKI (67) and nephrologist adjudication using RIFLE criteria (68)
Sensitivity	90.4% 'accuracy'#	Not calculated	100 %	91.2% (ICD-10 coded) and 90.5% (nephrologist adjudication)
Specificity		Not calculated	92.7 %	Not calculated

Table 1. 2 Summary of published AKI alert detection algorithms used in routine information health systems highlighting major similarities, differences, and performance (when quantified)

This table is by no means exhaustive but underscores the myriad ways in which even consensus AKI criteria, often amended, have been applied to automated detection. # The metric of accuracy was not defined. *The higher alert by AKIN or RIFLE criteria was taken as final. AKIN and RIFLE essentially equate to KDIGO criteria. § Alerts from people with KFRT may be suppressed locally at the laboratory or once received by the UK Renal Registry. Abbreviations: **MDRD**, Modification of Diet in Renal Disease; **eGFR**, estimated glomerular filtration rate, **RIFLE**, Risk, Failure, Injury, Loss, End-stage kidney disease; **AKIN**, Acute Kidney Injury Network; **KF**, kidney failure; **NHS**, National Health Service, **KDIGO**: Kidney Disease: Improving Global Outcomes, **ICD-10**, International Classification of Diseases 10th revision.

1.4.2 Validation of alerts

The validation of algorithms is important since alerts are used in the clinical management of patients and alert data are used to elucidate the epidemiology and outcomes of AKI.(69) AKI may be misclassified as chronic, and vice versa, CKD may be misclassified as acute. Algorithm validation has therefore traditionally been performed manually using clinical adjudication which is appropriate since AKI is usually a clinical diagnosis.(70,71)

A single-centre study in England, which adopted an AKIN e-alert algorithm, observed 6,037 AKI alerts in 2,619 patients over a 9-month period.(70) Alerts were validated by manually reviewing baseline SCr values of patients who triggered an alert and every SCr value over three random 24-hour periods were reviewed for AKI. Only 4/1,702 patients who were randomly reviewed were falsely missed, 103/6,037 (1.7%) did not develop AKI, and 194/6,037 (3.2%) were misclassified as AKI instead of CKD.

Others have validated rule-based algorithms in the ICU.(71) Sensitivity was 88% and specificity 96% when algorithm-identified AKI was compared to episodes identified by the nephrologist (algorithm-nephrologist). Agreement between the algorithm and nephrologist was also high (Cohen's Kappa 0.84; 95% CI 0.78, 0.89).

The NHS England algorithm has been externally validated using clinical adjudication and ICD-10 coded AKI in the Scottish GLOMMS II cohort.(67) The authors found the algorithm to be sensitive (only missing 8.8% of ICD-10 coded AKI episodes). Sensitivity improved to 96.7% when the algorithm was amended to extend the baseline reference period to 3 years (rather than 7 days [KDIGO] or 365 days [NHS England algorithm]) and include the lowest SCr as

opposed to the median in the preceding 8 – 365 days, as used in the NHS England algorithm. This extended period and the use of the lowest SCr may lead to lower specificity in determining the baseline SCr because SCr may change significantly within 3 years. Such manual validation work is not practical when large numbers of alerts are generated, such as received by the United Kingdom Renal Registry (UKRR).(72)

1.4.3 The United Kingdom Renal Registry

This thesis used data collected by the UKRR, detailed further here. The UKRR was established in 1995; a division of the UK Kidney Association (ukkidney.org). All 83 kidney centres in the UK are mandated by the NHS to submit data about adult and paediatric kidney patients with kidney failure requiring replacement therapy (KFRT), and more recently, infections and advanced CKD (eGFR < 30ml/min per 1.73m^2) not requiring KRT or receiving conservative care. The data are used for quality-of-care assessment, audit, and for research purposes. The UKRR has also been leveraged to conduct Registry-based clinical trials.(73)

For CKD and KF, data acquisitions are automatically collected through kidney centres' information technology systems. Centres in England, Wales and Northern Ireland submit data directly to the UKRR while the Scottish Renal Registry submits on behalf of its centres. Laboratory data are received by the centre before submission to the UKRR. Recently, the UKRR has also been able to link with routinely collected Hospital Episode Statistics (England), Patient Episode Database for Wales and Public Health England infections databases, improving data completeness.(74)

For AKI, NHS England issued a Patient Safety Alert in 2014 (NHS/PSA/D/2014/010) to all pathology providers to standardise AKI detection.(63) From March 2015, NHS England has required all 190 laboratories in England to integrate an algorithm to detect cases of AKI within its laboratory information management system (LIMS). Of 190, 39 laboratories submit data indirectly via an associated main laboratory. Data are not processed by the centre, unlike other non-AKI data, and are sent directly to the UKRR.

This falls under a wider 'Think Kidneys' initiative instituted by the Renal Association and UKRR (now the UK Kidney Association) which seeks to improve the quality of care and outcomes of patients with kidney disease (thinkkidneys.nhs.uk). Complementary resources to manage and prevent AKI were promoted under the second Patient Safety Alert issued in 2016.

The first UKRR report of AKI from 2018 data reported 1,524,398 alerts representing 488,856 patients received from 87.4% of laboratories in England (164 of 190) submitting suitable quality data.(72)

1.4.4 NHS England acute kidney injury detection algorithm

SCr results collected from primary and secondary care that are compatible with KDIGO criteria for AKI, fire an alert and are communicated to the requesting clinician in real-time.(11) Once a month, the laboratory sends these alerts, associated SCr results, and SCr results from the preceding 15-months to the UKRR.(72) Values for the subsequent 15-months after the alert are reported to the UKRR in the months following the alert. Alert and SCr data are linked by a unique Master Patient Index (MPI) using the pervasive NHS number. These data may then

be used to describe the incidence and outcomes of AKI across England, bench-mark Trusts within the country and serve as a quality improvement monitoring mechanism.

Laboratories may serve several Trusts, but not every Trust has a dedicated laboratory, and some may send samples to more than one laboratory. The disadvantage of this is that rates of AKI cannot be summarised by Trust and previous SCr data may be inaccessible by other laboratories outside of its network.

NHS number
Patient postcode
The AKI warning stage (based on creatinine KDIGO stage 1, 2 or 3; but not '0')
SCr result at alert
eGFR by MDRD and CKD-EPI equations
Identifiable and demographic information
Laboratory code
Location code (community versus hospital)
The SCr values from the preceding 15 months (separate 'creatinine' dataset)
The SCr values for the subsequent 15 months (separate 'creatinine' dataset)
Unique specimen number

Box 1. 1 Information the UKRR receives from laboratories once an alert is generated

Abbreviations**: NHS**, National Health Service; **AKI**, acute kidney injury; **KDIGO**, Kidney Disease: Improving Global Outcomes; **SCr**, serum creatinine; **eGFR**, estimated glomerular filtration rate; **MDRD**, Modification of Diet in Renal Disease; **CKD-EPI**, Chronic Kidney Disease Epidemiology study.

A flowchart of the algorithm may be found in Appendix 1, also available at https://www.england.nhs.uk/akiprogramme/aki-algorithm/. If a previous SCr result exists in the 365 days prior to the index creatinine (called 'C1'), the lowest value is used as the reference baseline if the result is within 0 – 7 days ('RV1') or the median of creatinine values within 8 – 365 days ('RV2'). This reference value (RV) is compared to C1. The highest ratio is preferred. If the ratio (C1/RV) is < 1.5, and the difference between values within 48 hours is < 26 µmol/L, no alert (called an 'AKI warning test') is generated. Otherwise, a ratio ≥1.5 (or absolute value > 26 µmol/L; KDIGO 1), ≥ 2.0 but < 3.0 (KDIGO 2) or ≥ 3.0 (or absolute result > 354 µmol/L; KDIGO 3) generates an AKI alert.(63)

If no prior SCr result can be found in the laboratory information management system (LIMS), which is limited to that laboratory; SCr results from other laboratories cannot be 'seen', C1 is compared to the laboratory reference interval, and if high, flagged as possible AKI or CKD with a recommendation to repeat testing. The UKRR does not receive these warning flags. The population reference interval may not be the same across all laboratories.

The algorithm discontinues generating an alert once these criteria are not met, but there is no definition of recovery included in the algorithm to delineate one AKI episode from another. Multiple alerts are generated per patient depending on how many SCr tests are requested (sometimes multiple per day) such that each patient may have multiple alerts during the same AKI episode. The UKRR does not receive the reason for the alert i.e., the threshold reached, or the specific RV (RV1 or RV2) used to generate the alert although pre-alert data are available.(63)

1.4.5 Data quality assessment

Nearly 99% of laboratories (189/190) across England are now sending AKI alert data.(72)

Alert data-files that are received are quality checked by the UKRR data management team. A classification system based on the completeness of key variables (NHS number, sex, postcode, and AKI stage) was adopted.(72) Date of birth has recently been added as well. Monthly completeness ratings are published on thinkkidneys.nhs.uk/aki/aki-data. Missing data items are reconciled with NHS Digital. The UKRR also attempts to recover missing data or errors directly from laboratories.

Some alert data for certain months may therefore be missing if classified as 'Red' (Table 1. 3). Data errors, corrupt data-files and failure to reconcile NHS numbers are reasons for missing data. Although data may have been missing because the laboratory had yet to implement the detection algorithm, once implemented, laboratories do consistently send alert data. Data are also reconciled with information i.e., demographics held by the NHS. In 2018, 87.4% of laboratories sent 'Green' or 'Amber' scored data(72)

Rating	Monthly data completeness	Yearly data completeness
Red	<50% complete – data removed	<6 months with data – laboratory removed
Amber	50 – 90% complete	12 months with 50 – 90% complete 6 – 9 months with >90% complete ≥9 months with 50 – 90% complete
Green	>90% complete	12 months with >90% complete ≥9 months with >90% complete

Table 1. 3 Quality grading assigned by the UKRR for the data it receives

Red is most incomplete; Green is most complete and Amber in-between (**RAG**). From the UK Renal Registry AKI 2018 report.

Only a proportion of laboratories currently send pre- and post-alert creatinine data as well. Pre-alert data are important because these should include the baseline SCr value that the current alert SCr value is compared with.(11)

How each laboratory has implemented this algorithm in its local LIMS is unknown and to date no work has been undertaken to examine this. Since these alerts are being transmitted to clinicians caring for these patients, being used to compare AKI rates between Trusts and emulated by others in other settings, it is important to ensure the algorithm is working accurately and consistently.

1.4.6 Routine health data collection in the Western Cape

Routine health data collected during the provision of health care in the Western Cape, South Africa, were also utilised in this thesis to describe the clinical epidemiology of kidney disease in the region of the City of Cape Town.

1.4.7 Setting: Healthcare provision and structure in South Africa

Provision of healthcare in South Africa is currently delivered by government-funded national and provincial government (public service), and by self-paid private medical insurance schemes. The majority of people utilizes the public service (71.4% of South Africans reported that they access public healthcare facilities in the 2016 household census), whereas it is estimated that ~85% of healthcare expenditure is consumed by the private health care sector.(75) Public healthcare is available for all, including legal citizens, refugees and asylum seekers. People who are unemployed and on social grants do not have to pay to access care. Otherwise, a tiered fee is charged for those who have the means, based on their income.

Levels of public healthcare include district based primary and secondary care, and regional and tertiary centres situated only in larger cities. For various reasons, not all South Africans (92.6% by self-report) present to the nearest healthcare facility when they fall ill, however, and it is not known how far people are willing to travel to access their preferred facility. In addition, healthcare seeking behaviours and barriers to accessing healthcare mean that not all members of the general population regularly attend government facilities. These factors make determining a denominator population using these data challenging when calculating disease rates and is likely to bias estimated incidence rates.(75)

1.4.8 Public health sector geography of the Western Cape

One of 9 provinces (Figure 1. 4), the Western Cape consists of an estimated 6.8 million people of the estimated total of 58.7 million living in South Africa in 2019.(76) In the year 2019 there were 14.2 million primary health care encounters, 288,199 patients admitted to district-level, 125,976 to regional-level and 140,392 to tertiary-level hospitals in the public sector.

There are two public tertiary level hospitals (excluding children's, psychiatric and tuberculosis, hospitals) in the Western Cape.(77) One of them, Groote Schuur Hospital also provides some specialist services (organ harvesting and transplant, adult endocrinology and cardiology) to the Eastern Cape province.

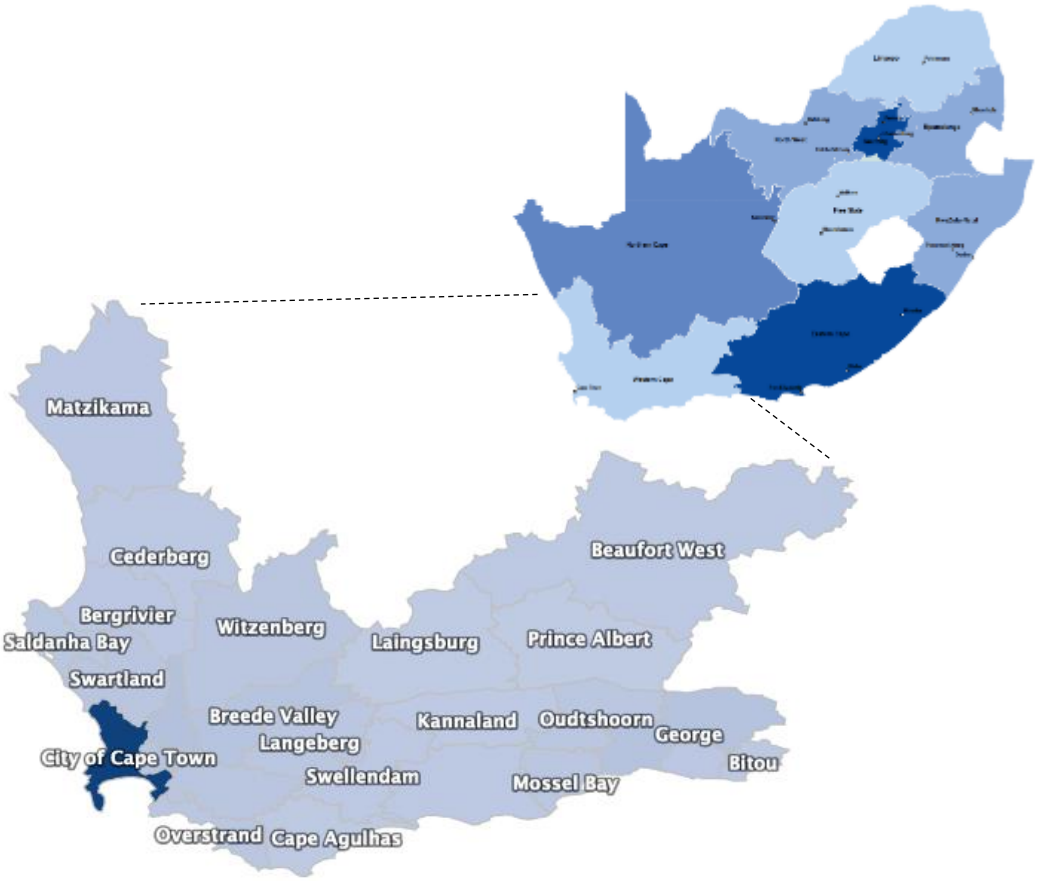


Figure 1. 4 Map of the Western Cape province (map of South Africa inset)

Districts of the Western Cape are labelled. The City of Cape Town, in dark blue, is the focus of parts this thesis. Re-printed from Statistics South Africa, statssa.gov.za.

1.4.9 Demographics and health indicators

The national population is estimated to be of majority black race (80.7%), followed by 8.8% mixed ancestry, 7.9% white and 2.6% Indian/Asian origin.(75)

In the Western Cape, life expectancy for men (66.8 years) and women (71.8 years) is the highest in the country. There were ~48,000 deaths in the Western Cape recorded in 2016. (78) Premature death was attributable to cardiovascular causes (19.8%), cancers (19.3%), other non-communicable diseases (17.1%), HIV and TB (15.3%) and injury (13.3%). Just under 280,000 people were on antiretroviral medication in 2018/2019.

1.4.10 The provincial Health Data Centre

The Western Cape Provincial Health Data Centre (PHDC), a division of the Western Cape Provincial Department of Health, was established in 2015 and receives individual-level electronic data from various recently established and legacy clinical and administrative source systems across all levels of public sector healthcare.(79) Deterministic and probabilistic record linkage using a Patient Master Index (PMI) unique identifier and other identifiers in the case of duplicates, matches all patient records to individuals within the Province of the Western Cape. Data are usually automatically uploaded from sources daily or periodically (disease registers) to construct a detailed longitudinal dataset within this health information exchange (HIE).

Hospital and clinic administrative systems

Laboratory (National Health Laboratory Service)

Pharmacy and dispensing

Hospital discharge summaries

Tuberculosis registers

HIV highly active antiretroviral therapy registers

Perinatal reports

Child Health reports

Radiology

Emergency centre tracking and hospital transfer (ambulance) services

Box 1. 2 Electronic data sources curated by the PHDC

These sources are from public healthcare facilities only.

Person, place (healthcare facility) and time are used to identify contact by patients ('encounters') through attendance to outpatient areas, medication dispensing, sampling for laboratory tests or admission to hospital.

Health conditions (such as HIV, diabetes, hypertension, tuberculosis, AKI and CKD) are inferred indirectly by 'common concepts' supported secondarily by administrative ICD-10 coded data. Common concepts are groupings of laboratory, medication (Anatomical Therapeutic Chemical coding), hospital visit and other data that are used to form inferred conditions. Visits to the kidney clinic, certain immunosuppressive medication and immunological blood tests are for example used to identify someone with a kidney transplant. Inferences are ascribed levels of certainty and only high certainty episodes using multiple evidence are used. Episodes for each inferred condition are defined for individual healthcare users, identifying episode start and end dates.

Encounters are linked with episodes, medication records, laboratory test results, comorbidities and outcomes (e.g., mortality) to form a cascade which can then be processed and accessed by healthcare practitioners through a clinical portal (Single Patient Viewer [SPV] electronic health record in pilot phase) or can be used to generate operational surveillance reports and customised extracts for use by the health service to support continuity of care.(79,80) Whereas encounter and episode data are stored in long datasets that contain multiple rows of data about each occurrence, cascades synthesise the data into a wide dataset containing fields of clinical relevance, with one data line per patient. The data fields presented in cascades are determined in collaboration with specialist clinicians who advise on the most clinically useful summary metrics for that particular episode.

Since communication between primary, secondary and tertiary care is fragmented, and digital health platforms generally restrict data access for the facility in which data was collected, these tools additionally allow for improved information sharing and clinical summaries through the SPV via web-based access, in order to support continuity of care.

Condition	Laboratory data types	Treatment data types	Diagnostic codes	Encounter evidence
HIV	ELISA, CD4, HIVVL parameters	HAART	ICD-10	ART clinic attendance
Diabetes	HbA1c > 6.5%, FG > 7.0 mmol/L, random or OGTT > 11.1 mmol/L	Antidiabetic medication	ICD-10	
Hypertension		Antihypertensive medication (e.g., hydrochlorothiazide)	ICD-10	
Chronic kidney disease				
Chronic kidney disease (non-KRT)	eGFR < 60 ml/min/1.73m^2 consecutively, 90 days apart			
Dialysis	Tissue typing		Dialysis procedure code, Renal ICD-10 N18.x	Kidney clinic attendance
Kidney transplantation	Tissue typing	Transplant medication (antithymocyte globulin, tacrolimus, cyclosporine, mycophenolate mofetil)	Transplantation procedure code	Kidney clinic attendance
Acute kidney injury				
Acute kidney injury	SCr > 100umol/L and previously < 100 µmol/L within last 90 d			

Table 1. 4 Data types used to infer common medical conditions

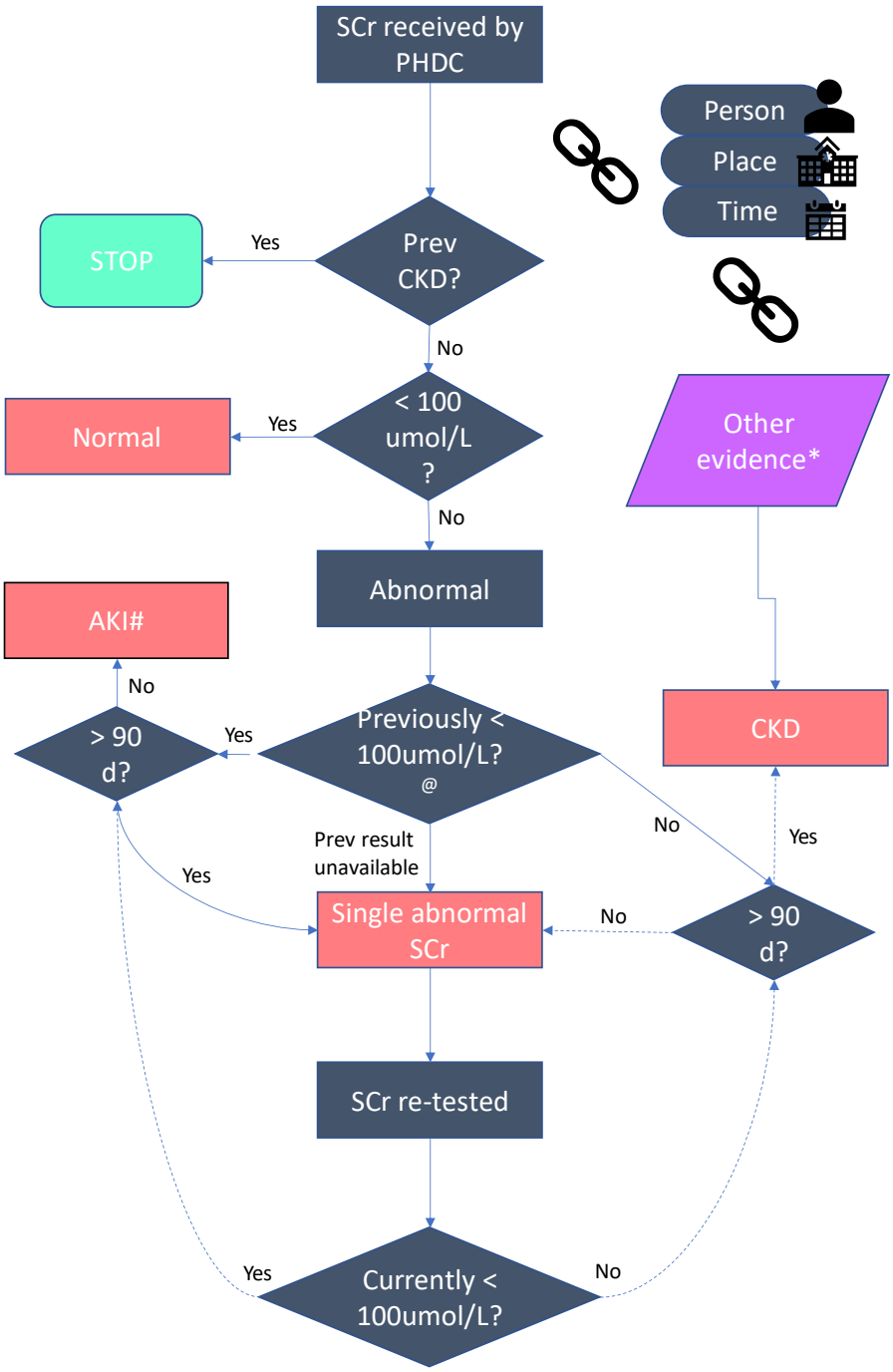
Complex algorithms using several data types are used to infer medical conditions. Because ICD-10 codes are often insensitive, frequency and patterns of healthcare facility engagement, laboratory results and dispensed medications are used as evidence of morbidity. Evidence is scored and weighted such that certain combinations of evidence are attributed as 'high evidence'. Abbreviations: **ELISA**, enzyme-linked immunosorbent assay; **HAART**, highly active antiretroviral treatment; **ART**, antiretroviral treatment; **HIV**, human immunodeficiency virus; **HIVVL**, HIV viral load; CD4, cluster of differentiation 4; **FG**, fasting glucose; **OGTT**, oral glucose tolerance test; **HbA1c**, glycated haemoglobin; **eGFR**, estimate glomerular filtration rate; **SCr**, serum creatinine.

1.4.11 Provincial Health Data Centre kidney disease cascade

Abnormal SCr results are assigned one of four possibilities using a pragmatic algorithm, summarised in Figure 1. 5:

1. Normal (SCr < 100 µmol/L)
2. 'Single abnormal creatinine' if the SCr is > 100 µmol/L but no preceding value exists to compare it to,
3. AKI if >100 µmol/L but previously <100 µmol/L,
4. and CKD (Table 1. 4).

A 'single abnormal creatinine' may later be re-assigned to AKI or CKD once further SCr/ eGFR results are available. AKI and CKD episodes are treated by the PHDC as mutually exclusive such that once a patient has an inferred diagnosis of CKD they cannot be subsequently flagged as having AKI after the CKD ascertainment date, regardless of subsequent SCr measurements.



*** Other evidence of CKD include KRT initiation, ICD-10 codes**
@ The look-back time period is indefinite
 # AKI episode terminates once subsequent SCr< 100 umol/L or 90 days elapse
This symbol denotes data linkage

Figure 1. 5 Kidney disease cascade used by the PHDC to identify patients with AKI and CKD

The dashed line refers to the later re-assignment of a single abnormal creatinine. Abbreviations: **SCr**, serum creatinine; **PHDC**, Provincial Health Data Centre; **AKI**, acute kidney injury; **CKD**, chronic kidney disease; **KRT**, kidney replacement therapy.

The PHDC offers a unique opportunity to explore the clinical epidemiology of kidney disease in a South African setting by characterising the frequency, outcomes and healthcare utilisation of people with kidney disease using routinely collected data. Although source systems may be the same or similar in other areas, no comparable African data centre exists outside of the Western Cape Province in South Africa. It has played an integral part in the public health response to the Coronavirus-19 pandemic (www.coronavirus.westerncape.gov.za/covid-19-dashboard).

AKI does not always resolve regardless of how precisely or early it is detected, and CKD may result with serious implications for health.

1.5 EPIDEMIOLOGY OF CHRONIC KIDNEY DISEASE

Incidence/prevalence estimates of CKD are dependent on how the glomerular filtration rate is measured (or estimated), if albuminuria criteria are included and if enough time is given between GFRs to ensure it represents steady-state.(81) Calculated GFR using creatinine-based equations under- or over-estimate the measured GFR and might under- or over-estimate CKD. For these and other sampling reasons, there is wide variability in the reported incidence of CKD.(81)

1.5.1 World-wide prevalence of chronic kidney disease

In the latest Global Burden of Disease report in 2017 representing 195 countries, the number people worldwide with CKD was estimated to be around 698 million, with 32 million attributable to type 1 diabetes, 126 million attributable to type 2 diabetes, 24 million due to hypertension and 28 million to glomerulonephritis.(82) These numbers are expected to rise with the ever-ageing population and increasing incidence of diabetes according to predictions by the United Nations.(83)

In a meta-analysis by Hill et al, nearly 7 million patients mostly from Europe, USA and Asia were analysed using MDRD and CKD-EPI estimated GFR equations.(84) The proportion of studied individuals with evidence of CKD was 13.4% (95% CI 11.72, 15.14; I^2 99%) with most falling into stage 3 (7.6%), whilst the proportion of patients with advanced disease was low (0.4% stage 4 and 0.1% stage 5). The highest percentage was amongst females (14.6% vs 12.8% for males), increasing age (27.6% and 34.3% in the 60 – 69- and 70 – 79-year-old age groups, respectively), and patients with hypertension and diabetes.

1.5.2 Chronic kidney disease in Africa

The prevalence proportion of people with CKD was estimated to be only slightly higher in Africa compared to world-wide estimates (15.8%; 95% CI 12.1, 19.9) in a meta-analysis of just over 98,000 adults for all stages of CKD with few having more advanced disease (4.6% stages 3 – 5).(85) In contrast to other continents, the median age was 43 years old. HIV (27.3%), hypertension (35.6%) and diabetes (32.6%) were frequent comorbidities. Only 22/54 African countries were represented. The sampled populations were varied (general population versus hospital based), the GFR was mostly estimated by MDRD, and Cockcroft-Gault equations and proteinuria quantified using the urine dipstick. Compared to mGFR, there is a significant risk that measures of CKD frequency are miscalculated because the eGFR is grossly underestimated in Africa using SCr based equations with or without the ethnic correction factor.(86) There are also other methodological factors which may explain why differences were observed between world-wide and estimates from Africa. Researchers would often choose to use convenience sampling of medical facilities where there are established research capabilities, personal contacts and academic centres. The results may over-represent centres where kidney disease is more likely to be recognised, especially if nephrologists were responsible for screening patients for kidney disease. In Africa, where specialist nephrology services are more scarce and researchers are only likely to be in contact with academic centres in urbanised cities, referred patients are likely to have advanced kidney disease and only referred for dialysis services.(87,88)

1.5.3 CKD progression

The European CKD Burden Consortium showed that annualised kidney function decline was quite variable between countries, ranging from 0.77 (95% CI 0.45, 1.08) in the Belgian cohort to 2.43 (95% CI 2.11, 2.75) ml/min/1.73m^2 in the Spanish cohort. This is interesting since the baseline eGFR was lowest in the Belgian cohort (mean 19.9 ml/min/1.73m^2 [SD 5.4]) and highest for the Spanish cohort (mean 37.7 ml/min/1.73m^2 [SD 11.5]). This may be explained by other important patient related clinical (including patient exclusion criteria) or country specific factors, such as referral to nephrology services, management of risk factors, and KRT initiation practices.(36)

1.5.4 Risk factors for chronic kidney disease progression

AKI has recently not only been recognised as a strong risk factor for the development of CKD, but also for the progression to KF and death.(51,89) A meta-analysis performed by Coca et al found that the pooled rate of development of *de novo* CKD and KF following AKI was 25.8 per 100 person-years and 8.6 per 100 person-years, respectively. (47) The risk of developing CKD and progressing to KF increased as AKI severity increased, but of importance, was still increased for 'mild' AKI (pooled adjusted HR 2.0 [95% CI 1.4, 2.8] for CKD and HR 2.3 [95% CI 1.7, 2.3] for KF).

Other risk factors for CKD progression, as confirmed in a meta-analysis, include male sex (HR 1.37; 95% CI 1.17, 1.62) and proteinuria >1g/day (HR 1.64; 95% CI 1.01, 2.66) for the development of KF from stages 3 – 5 CKD.(90) The Assessment, Serial Evaluation, and Subsequent Sequalae in AKI (ASSESS-AKI) Study found that for each doubling in urine ACR level measured 3 months following AKI, there was a 1.53 times increase in HR (95% CI 1.45,

1.62) associated with halving of eGFR or development of KF.(91) Furthermore, in the Kidney Failure Risk Equation (KFRE), developed and validated in two Canadian cohorts with stages 3 – 5 CKD, included male sex, albuminuria, eGFR, and age (KFRE-4) were predictors of reaching kidney failure faster at 2 and 5 years. The authors also concluded that serum markers of bone turnover (calcium, phosphate), acidosis (bicarbonate) and inflammation/ malnutrition (albumin) were also predictive (KFRE-8), but less practical because of infrequent measurement of these parameters.(92)

In addition, the CKD-PC identified systolic blood pressure, a history of cardiovascular disease, black race, and diabetes as risk factors for KRT initiation.(93) They also found that age was protective for KRT initiation. This may be explained by the competing risk[5] of death; corroborated in a US Veterans Affairs study: older patients (≥ 75 years old) with advanced CKD (eGFR 15 – 30 ml/min per 1.73m^2) tended to progress to KF slower than younger patients, but had a higher risk of dying than developing KF. Elderly patients may decide on a more conservative approach to KF management, because of increased mortality associated with initiating dialysis.(94,95)

1.5.5 CKD progression and mortality

Not only is mortality a competing risk for the development of KF, and in many cases the mortality risk is higher than the risk of developing KF, but progression is also a risk factor for death.(96) Although rapid decline (> –5ml/min/1.73m^2 per year) in eGFR was uncommon (11%) in a CKD-PC meta-analysis, all-cause mortality was higher for people experiencing eGFR decline of –6ml/min/1.73m^2 compared to 0ml/min/1.73m^2 (adjusted HR 1.25; 95% CI 1.09,

[5] Competing risks will be explained in more detail in section 8.1.3.

1.44 for the pooled CKD cohorts and 1.15; 95% CI 1.01, 1.31 for the pooled general population cohorts). The relationship appeared to be U-shaped i.e., similar for people with +6ml/min/1.73m^2, potentially because of limitations of eGFR based on creatinine.(97) Risk was similar for cardiovascular and non-cardiovascular causes of death.

Studies have mostly evaluated traditional risk factors for progression and death. Understanding non-traditional risk may also inform biological (dys)function responsible for progression.

1.6 PROTEIN ASSOCIATIONS WITH CHRONIC KIDNEY DISEASE PROGRESSION

eGFR, as a proxy of glomerular function, and urinary albumin, as a proxy of glomerular, tubulo-interstitial, and endothelial integrity, do not reflect the other functions of the kidney, such as tubular, interstitial, and vascular activities. Following an insult, gene transcription, which through a series of biological processes (translation) results in synthesis of proteins that have functions to protect and repair. Collateral damage to normal tissue may occur when the response is exaggerated, however, and is responsible for perturbations in normal kidney functioning.(98) Proteomic analysis, the study of proteins, of urine and blood, that have been associated with glomerular, tubular injury and repair, and other pathophysiological processes, may mechanistically explain why some individuals more rapidly experience decline in their kidney function compared to others and serve as biomarkers or novel predictors of CKD progression.

Although the urinary proteome seems like the most obvious means to measure proteins, since the healthy kidney reabsorbs nearly all lost protein and the loss of proteins is the direct result of kidney damage, the abundance of protein in the urine of unhealthy kidneys may simply indicate a defective sieving or reabsorption mechanism and not a direct explanatory link to the damage itself.(99,100) Proteomic analysis of the blood therefore may be a better indicator of the inciting pathophysiology, although some serum proteins may be increased simply because of reduced kidney filtration function.(101)

Grams et al. for example identified 13 out of 4,877 widely expressed proteins, representing inflammation and other metabolic processes, associated with the development of KF or 50% decline in eGFR in people without CKD in the Atherosclerosis In Communities (ARIC) Study. This was externally validated in two CKD cohorts: The Chronic Renal Insufficiency Cohort (CRIC) Study and African American Study of Kidney Diseases Study (AASK) making the biological associations highly generalisable to both people with and without CKD.(102)

Multiplex proteomic assays, such as developed by Olink® (www.olink.com, Uppsala, Sweden), identified 20 proteins associated with kidney function decline (eGFR slope per year) using a discovery cohort derived from the Prospective Investigation of the Vasculature in Uppsala Seniors Study (PIVUS) and replicated in the Uppsala Longitudinal Study of Adult Men (ULSAM).(103) In this study, however, there were limitations in the study population (a single region in Sweden with normal to stage 1 – 3 CKD and replicated only in men), and simple linear regression was used to calculate eGFR slope which ignores important methodological considerations.

1.6.1 Biological mechanisms

The biological mechanisms of progression to KF have been hypothesised to result from (a) low nephron endowment at birth (for various reasons related to maternal and in utero conditions), (b) injury to the podocyte by primary and secondary causes, (c) direct nephrotoxicity of proteinuria, and (d) scarring and fibrosis of the nephron and interstitium in addition to local haemodynamic changes.(100,104–106)

The tubule is susceptible to toxic, ischaemic and inflammatory insults. In the Systolic Blood Pressure Intervention Trial (SPRINT), tubular biomarkers were assayed. It was found that eGFR increased by 0.34% (95% CI 0.08, 0.60) per year for every doubling in urinary uromodulin and eGFR decreased by –0.10% (95% CI –0.18, –0.02) with elevated urinary β-2 microglobulin in the standard blood pressure treatment arm, but the association did not remain in the intensive blood pressure treatment arm.(107) This may be because a lower blood pressure offered better kidney protection against tubular toxicity or support that these non-traditional biomarkers are more important than traditional risk factors.(100)

Further scrutiny reveals that biological processes such as inflammation, dysregulated immune responses, and disordered metabolism underpin traditional risk factors for cardiovascular disease and CKD progression.

1.6.2 Inflammation

Amdur et al. investigated the relationship between interleukin-1, interleukin-1 receptor antagonist, interleukin-6, tumour-necrosis factor α, tissue growth factor β, high-sensitivity C-reactive protein (CRP), fibrinogen, and albumin with CKD progression, and separately, a

composite of atherosclerotic cardiovascular events and death, in the CRIC cohort.(98,108) They found that increases in these cytokines and acute-phase proteins, and decrease in albumin were associated with faster eGFR decline and atherosclerotic cardiovascular disease-death composite.

1.6.3 Thrombosis

CKD is a prothrombotic state which is exacerbated by inflammation.(109,110) Given that the kidney, especially the medulla, is sensitive to hypoxia, disturbances in the microcirculation, from microthrombi caused by abnormal fibrin and platelet activation, adhesion, and fibrinolysis, may elicit nephron damage.(111,112) In the Rotterdam Study, a higher von Willebrand factor: ADAMTS-13 (a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13) ratio, which represents prothrombotic/ antithrombotic balance, was associated with a steeper eGFR slope (-0.06 [95% CI -0.09, -0.03] ml/min/1.73m^2 per year) and higher risk of halving of eGFR (OR 1.40; 95% CI 1.02, 1.93).(113) These are small (slope) and wide (risk) estimates, but may be underestimated because albuminuria was not measured. Support for inflammatory (CRP, albumin) and pro-coagulant (fibrinogen, factor VII) mediators of progression were also found in the Cardiovascular Health Study of older (> 65 years) community dwellers.(114)

1.6.4 Oxidative stress

Advanced Glycation End-products (AGEs), which accumulate in uraemia and diabetes, have been implicated in CKD progression, cardiovascular disease, and worsening of inflammation and platelet hyperactivation.(112,115) Evidence from murine models showed that oxidative proteins cause podocyte damage via interactions with the receptor for AGEs (RAGE) eliciting

injurious oxygen free radicals.(116) Podocyte loss results in proteinuria, which itself is tubulo-toxic and is associated with venous thromboembolism.(117,118) Also, albuminuria has independently been linked to the development of cardiovascular disease, highlighting the global endothelial dysfunction[6] that occurs throughout the vasculature, especially in diabetes, that is associated with kidney and cardiovascular disease.(119–121)

A pro-inflammatory, pro-thrombotic environment and oxidative stress culminates in disturbed homeostasis leading to thrombosis, atherosclerosis, arterial stiffening and fibrosis affecting the micro- and macro-vasculature more generally and kidney and cardiac structures specifically.(122,123) This phenotypically manifests as ischaemic heart disease, cerebrovascular disease, peripheral vascular disease, and progression to kidney and heart failure.

1.6.5 CKD progression and cardiovascular disease

Cardiac proteins may accumulate in CKD simply because of decreased kidney excretion but accumulation may also result from increased production associated with structural heart disease and kidney-heart crosstalk such as in the cardiorenal syndromes.(124,125) For example, kidney perfusion may decrease because of heart failure by causing kidney congestion, and vascular atherosclerosis and calcification of the heart, accelerated by uraemia, may precipitate or worsen cardiac ischaemia.(126–128)

[6] The endothelium is the inner most cellular layer of all blood vessels. It interacts with the cells below it within the vessel wall and proteins that lie on its luminal surface (glycocalyx layer).

A CRIC sub-study highlighted the important association of cardiac dysfunction with CKD progression: N-terminal pro-B-type natriuretic peptide (NT-proBNP), a myocardial stretch protein, and Growth-Differentiating Factor 15 (GDF-15), implicated in cardiac inflammation and remodelling, were associated with 50% decline in eGFR or need for KRT initiation.(129) High-sensitivity troponin T (hs-TnT), which is released by myocytes during cardiac injury including ischaemia, was only associated with progression in people with pre-existing CVD, although people with pre-existing heart failure were excluded, but others (ARIC Study) have found that hs-TnT (as well as NT-proBNP) to predict progression.(130) In a related CRIC sub-study, these biomarkers also associated with all-cause and CV mortality, underscoring the biological relationship between CKD progression and non-traditional cardiovascular risk.(131)

It is clear that not only is there an epidemiological link between CKD and cardiovascular disease, but a biological one too, protein relationships are, however, complex.

1.6.6 Clustering proteins into biological systems

Proteins do not act in isolation: they function in coordinated biological pathways.(132) Proteins may exert their effect proximally or distally within a pathway, and therefore a small concentration of a proximal protein may have exponential distal effects. Within the system, proteins may upregulate or downregulate other proteins, or have positive or negative feedback on other, or their own processes. This establishes a delicate balance between pro- and anti- inflammatory, thrombotic, fibrotic, oxidant and other processes. Kidney injury may result if these processes are dysregulated.(98) Kidney function is particularly complex because there are several cell types that constitute the nephron and supporting structures in which intra- and inter-cellular signalling occur.

The SPRINT Trial categorised 10 biomarkers into 4 clusters representing tubule injury/ repair, tubule injury/ fibrosis, tubule resorption, and tubule reserve/ mineral metabolism using exploratory factor analysis.(132) Clusters were associated with –0.06% (95% CI –0.24, 0.12%), –0.16% (95% CI –0.33, –0.01%), –0.07% (95% CI –0.25, –0.12), and –0.58% (95% CI –0.39, –0.67%) decline in eGFR per year for each 1 SD increase in composite score, respectively. This may suggest that particular regulatory pathways influence CKD progression more than others. Using a statistical approach to protein clustering, however, ignores the complex biological interaction that different proteins have within a pathway or between each other.

1.6.7 Gene ontology

The Gene Ontology (GO) Consortium introduced a standardised organisation of different components of a biological system and has established standardised vocabulary (GO terms) for defining the relationship between gene products (e.g., proteins and transcription factors) found in thousands of species including *homo sapiens*.(133,134) The GO are represented as directed acyclic graphs (DAGs) where each node is a particular component of the broader biological system. Hierarchal GO terms, from parent (high-level) to child (low-level), are structurally arranged and catalogued into *biological processes* (e.g., extracellular matrix formation) which consist of multiple *molecular functions* (e.g., protein binding) within *cellular components* (e.g., the extracellular space). Each component of the above organisation is annotated with an identifier which may be common across different, but ostensibly inter-connected, biological pathways and even organisms. The *extracellular matrix organisation* pathway is demonstrated in Figure 1. 6. Knowledge is curated automatically and manually using various levels of evidence (experimental, non-experimental, automatic [non-human

curation]). GO terms are used to annotate genes, transcripts and proteins, thus providing a machine-readable annotation system to catalogue the functioning of molecular species. Searchable web-based platforms and standalone bioinformatic software make GO catalogues accessible and allow proteins to be grouped into modules that represent specific biological functions, interactions, and pathways.(135–137)

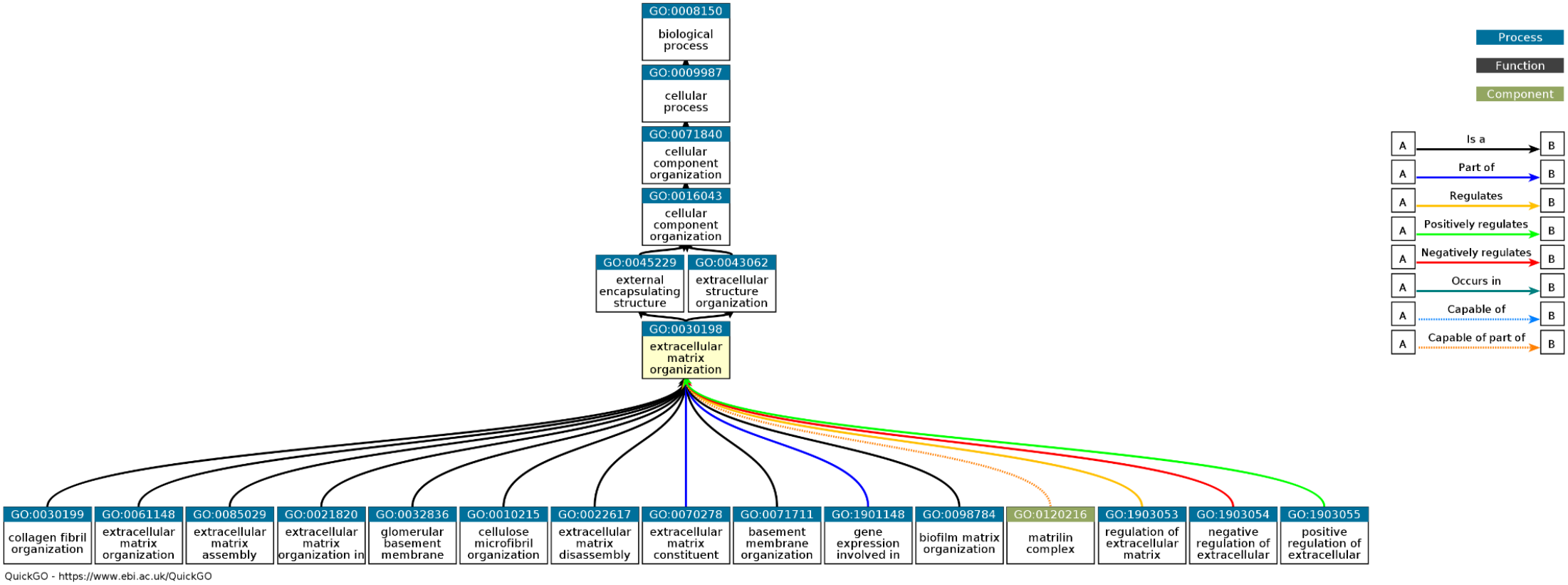


Figure 1. 6 Gene ontology hierarchy for the extracellular matrix organisation pathway

Generated using QuickGO search engine.(138) The key is shown on the right.

Studies have thus far mostly explored the diagnostic, predictive and mechanistic utility of protein biomarkers in patients with mild-moderate CKD with the hope that progression may be mitigated early before KF develops.(139) There is an unmet need to study patients with already advanced disease, who without kidney and cardiac protective interventions are at highest risk of requiring KRT or dying from their kidney and cardiovascular disease.(93) In addition, studies of CKD progression have mostly used a binary outcome of development of KF or KRT initiation. Using changes in eGFR may be more informative in advanced disease when the eGFR is already low and, when combined with bioinformatics and proteomics, highlight which biological mechanisms are associated with faster kidney function decline.(98)

The current work used eGFR and proteomic analysis collected in the European Quality Study of treatment in advanced CKD (EQUAL) to explore the association of cardiovascular and cardiometabolic proteins with kidney function decline.

1.6.8 Description of the EQUAL Study

Older patients (>65 years old) with incident advanced chronic kidney disease (with at least one eGFR < 20ml/min/ 1.73 m^2 by MDRD in the previous 6 months) referred to the nephrology service were prospectively enrolled in a multi-national observational study in participating European study sites (the United Kingdom, Sweden, Italy, Poland, the Netherlands and Germany) to investigate the associations between timing of initiation of dialysis (or transplantation) and uraemic symptom severity, the decision-making process and quality of life.(140) During April 2012 through December 2018, 1,754 participants were recruited. Participants were followed up until death, discharge back to primary care or end of follow-up at 4 years.

Clinical details (demographics, lifestyle, comorbidities, cause of kidney disease, anthropometry, medications, and laboratory data) were collected at baseline and at follow up (6 monthly until eGFR < 10ml /min/ 1.73 m^2, thereafter 3 monthly). The decision to start dialysis was patient-nephrologist shared and no GFR threshold was pre-specified (non-interventional study).

Plasma samples were also collected at the baseline visit and appropriately stored for later proteomic analysis. The investigators selected two protein panels, each comprising 92 proteins, developed by the Olink® proteomics laboratory (www.olink.com, Uppsala, Sweden): Olink® target 96 Cardiovascular II and Olink® target 96 Cardiometabolic. The proteins included are involved in myriad biological processes including immune/ inflammatory, coagulation, angiogenesis, apoptosis, cell adhesion and intracellular signalling pathways. A list of the proteins analysed in this thesis are provided in Appendix 2.

1.6.9 Olink® technology

Aliquots were analysed using the Proximity Extension Assay (PEA) platform (Olink Proteomics). PEA, a novel proteomic technology, involves three steps:

1. two oligonucleotide-labelled antibodies specific to each protein are added as a reagent to the sample, which bind to the target antigen, if present,
2. the oligonucleotide-antibody pair hybridize when in close proximity to each other and extend in the presence of DNA polymerase,

3. resulting DNA 'barcodes' are then measured using standard real-time polymerase chain reaction (PCR) techniques after amplification (Fluidigm® BioMark™ High Definition System).(141)

Advantages of using this technology are that multiple (96 analytes) may be tested simultaneously, requires a much smaller sample (1 microlitre), may be used with serum, plasma, urine, tissue, and other bio-fluids, and compares well to single-plex technologies such as chemiluminescent and enzyme-linked (ELISA) immunoassay methods. Validation of assays has been conducted by Olink® (https://www.olink.com/application/development-and-validation-of-customized-pea-biomarker-panels/).

Protein levels are reported as Normalised Protein eXpression (NPX) units, normalised to the median of the assay across plates and batches, on a log-2 scale. Two different NPX levels for two different proteins therefore do not necessarily have the same absolute concentration but an increase of 1 NPX can be interpreted as a doubling of the absolute protein concentration. Limits of detection and lower and upper limits of quantification (in picograms per millilitre) as well as normal levels (median, 10^{th} and 90^{th} percentile NPX units) have been defined for each protein by the manufacturer. In addition to 6 external negative and positive controls, all samples are also spiked with 4 internal controls that verify each step (immunoreaction, extension, and amplification/ detection) and are used to determine the level of detection of each protein, which is plate and sample batch specific.

1.7 CHAPTER SUMMARY

The collection of large amounts of information from source systems accumulated during routine healthcare facility activities, electronic medical records used to monitor and manage patients, called the digitome, and from biological samples, such as the proteome, has led to the advent of 'big data' and systems biology which has increasingly been used to study disorders of the kidney.(32,70)

There are efficiency, study design, precision, and research dimensionality advantages to using big data to study the spectrum of potentially reversible AKI and irreversible CKD epidemiology.

Rule-based algorithms, based on consensus criteria for AKI, may be embedded in pre-existing laboratory or electronic health record systems. The logic and implementation of such algorithms need to be validated in the settings in which they are expected to operate.

Kidney disease may be identified, and progress monitored clinically, using traditional biomarkers (SCr) and methods (eGFR), but novel proteomic markers of glomerular and non-glomerular dysfunction, involving other compartments of the kidney and more broadly, the vasculature, have been increasingly studied to understand the biological basis of disordered function, including inflammation, oxidative stress and thrombosis, and progression to KF. Furthermore, it may be possible to capture modules of grouped proteins, using bioinformatics that are involved in inter-related pathways to identify specific biological processes that are most significantly associated with CKD progression.

The aims of this thesis were to investigate the implementation of the NHS England algorithm (using routinely collected data received by the UKKR), describe the clinical epidemiology of AKI, CKD and acute-on-chronic kidney disease in the City of Cape Town, Western Cape, South Africa (using routinely collected data by the PHDC) and explore the association of novel proteomic signatures with kidney function decline in the EQUAL Study.

1.8 RESEARCH OVERVIEW

This research is divided into three work packages. Since this was a cotutelle United Kingdom-South African university joint PhD, multiple secondary data sources from the UK, Europe and South Africa were accessed to address these aims.

1.8.1 Objectives and research questions

1. Validate the NHS England mandated AKI detection algorithm using routinely collected AKI surveillance data submitted to the UKRR.
    a. How consistently is the NHS England PSA algorithm implemented in laboratories sending AKI alerts to the UK Renal Registry?

2. Describe the frequency of kidney disease, clinical characteristics of patients developing kidney disease, and their kidney and vital outcomes in the City of Cape Town using routinely collected data from the Province of the Western Cape, South Africa.
    a. Acute kidney injury clinical epidemiology
    b. Compare the PHDC and NHSE AKI detection algorithms

c. Chronic kidney disease clinical epidemiology

d. Acute-on-chronic kidney disease clinical epidemiology

3. Investigate the association between cardiovascular and cardiometabolic proteins with the trajectory of kidney function decline in older patients with advanced chronic kidney disease accessing nephrology services in Europe.

    a. Which cardiometabolic and cardiovascular disease related proteins and biological pathways are associated with kidney function trajectory in older persons with advanced kidney disease?

2 THE METHODS USED TO INVESTIGATE THE CONSISTENCY AND IMPLEMENTATION OF THE NATIONAL HEALTH SERVICES ENGLAND ACUTE KIDNEY INJURY DETECTION ALGORITHM

2.1 INTRODUCTION

How consistently was the NHS England AKI detection algorithm implemented in laboratories sending AKI alerts to the UK renal registry? The UKRR and the data that it receives were introduced in section 1.4.3. This chapter outlines the methods that were used to answer the above research question.

2.2 METHODS

2.2.1 Study population

For this study, data from laboratories that submitted high-quality[7] alert and pre- and post-alert[8] SCr results from the date of implementation of the NHS England algorithm (1 December 2014) through 30 September 2020 were included (37/187 laboratories that currently submit data).

Individuals aged < 18 years old were excluded as the reference ranges for children are age specific and the definition of AKI is different. Other exclusions were necessary based on

[7] Alert data are classified by the UKRR as 'Green', 'Amber', or 'Red' depending on completeness. High quality is considered 'Green'. See section 1.4.5.
[8] The UKRR receives SCr data for 15 months before and after an alert.

missing/ irregular data and duplicates. Age > 99 were excluded because of concerns of incorrect recording of the year of birth.

2.2.2 Central alert development

Local laboratory-generated alerts were compared with alerts derived using logic that simulated the NHS England algorithm ('central code'). This central code was executed within the alert and pre- and post-alert SCr data received by the UKRR.(63) The central code was written by the candidate using Stata version 17 (StataCorp, TX, USA). A summary of how this was performed is provided in Box 2. 1. An index SCr at current testing was compared to historical baseline values for the same individual, if available, based on KDIGO creatinine criteria.(11,63) The NHS England algorithm, summarised in section 1.4.4 and shown in Appendix 1, deviates from KDIGO criteria in that the look-back period is extended to 365 days.(63) The central code was executed in the usage data, described above, to generate 'central alerts'. Central AKI 0, 1, 2 and 3 and warning flags were created. The date of the first-ever received alert was taken as the date on which the algorithm went live within that laboratory. Local laboratory alerts were received as AKI 1,2 or 3. AKI 0 alerts were inferred centrally if no alert was received for that particular SCr value. Only SCr values that occurred after the live date, and not associated with an AKI alert, were recoded as AKI 0 because the algorithm had to have been active to theoretically fire an alert, if appropriate.

Preparation

- The SCr dataset in long format was reformatted into wide format.
- The long dataset was merged with the wide dataset ('long-wide' dataset). This allows the index SCr to be compared with historical values in the same row of data.

The long-wide dataset was cloned into 3 identical datasets: 'long-wide', '7 d', '365 d', then:

Using the long-wide dataset

1. Previous results available in preceding 365 d were identified.
2. Flags were generated for index SCr (C1) results when baseline creatinine values were unavailable in the preceding 365 d. If the C1 was within the laboratory reference interval (RI) then flags 'no flag', lower than the lower limit of the RI 'low flag', or higher than the upper limit of the RI 'high flag' were produced.

Using the 7 d dataset

3. The lowest SCr within 0 – 7 d before the C1 was identified as the baseline (RV1). The RV1:C1 ratio was calculated.

Using the 365 d dataset

4. The median SCr within 8 – 365 d before the C1 was calculated as the baseline (RV2). The RV2:C1 ratio was computed.

Long-wide, 7 d and 365 d datasets were re-combined:

5. The ratios in steps 3 and 4 were compared and the highest was preferred (RV). The difference between C1 and RV was calculated.

Central alerts that were > 1.5 * RV and:

6. A SCr difference > 26 µmol/L within 48 h was identified and flagged as AKI stage 1[§]
7. A SCr difference > 354 µmol/L or ratio > 3 was flagged as AKI stage 3
8. A ratio between 2 – 3 was flagged as AKI stage 2
9. A ratio between 1.5 – 2 was flagged as AKI stage 1

The dataset was saved and merged with the original laboratory alert dataset.

Box 2. 1 Summary of Stata code to simulate the NHS England AKI alert detection algorithm

dataset cloning was helpful because of data manipulation. § If the SCr had risen >26 µmol/L but not within 48 h, a warning flag was generated. Stata do file available at GitHub Repository: https://github.com/RyAylwd/NHSEnglandAKIalgorithm. Abbreviations: **AKI**, acute kidney injury; **SCr**, serum creatinine; **C1**, index creatinine; **RV**, reference value i.e., baseline creatinine; **d**, day; **h**, hour.

2.2.3 Agreement assessment using inter-rater methods

Table 2. 1 shows a hypothetical example of pairs of laboratory alert and central alerts that were compared.

	AKI alerts by KDIGO stage		Agreement
Creatinine µmol/L	**Laboratory**	**Central**	
75	0	0	Perfect agreement
100	2	1	Partial agreement
150	3	1	Disagreement

Table 2. 1 Illustration of hypothetical pairs of laboratory and centrally-coded AKI alerts compared for the same creatinine value

This example demonstrates how alerts were compared. It would be expected that the laboratory and central alerts should always agree if the same SCr data are available, and the same mathematical logic were applied. Abbreviations: **AKI**, acute kidney injury; **KDIGO**, Kidney Diseases: Improving Global Outcomes; **SCr**, serum creatinine.

Methods first developed for use in education and subsequently adopted by the medical community were used. Cohen's Kappa was the first to be published.(142) The pair-wise concordance between individual sets of alerts was of interest. Inter-rater reliability methods were used to assess the tendency that the laboratory code and central code independently classified AKI stages in the same way.(142) Raters, for example two nephrologists, might guess or randomly choose a category rather than use objective evidence to classify AKI into a finite number of 0 – 3 stages and so there might be 'chance agreement'. All methods assess the expected (by chance) agreement and compare this to the observed agreement. The laboratory and central code being compared are not subjective *sapiens* but may disagree for other reasons. Firstly, the laboratory code may have been altered to suit local opinions or limitations of the LIMS. There was no oversight over the operationalization of the algorithm within the

LIMS or laboratories.(6) Secondly, the data that the UKRR receives may be missing SCr values that the central code would not have 'seen' resulting in different classifications.

A well described problem, called the Kappa paradox, underestimates agreement when the unit of comparison is unbalanced.(143) This might occur because the central code does not compute the same number of alerts as laboratories submitted. Gwet's AC1 (agreement coefficient) overcomes this problem and so was used to assess uniformity of alerts produced. Gwet's AC1 equation was solved: the number of times the laboratory and central code computed the same AKI stage was divided by the total number of alerts (A). From A, chance agreement (C) was subtracted and divided by 1 – C.(144) Percent positive agreement (PPA) was also calculated by dividing the number of times the laboratory and central code both staged the same AKI stage by the total number of alerts generated by the central code. The user-written `kappaetc` Stata command was used to calculate PPA and Gwet's AC1.(145) Agreement coefficient magnitude, which ranges from -1.00 to +1.00, was interpreted as recommended by Landis and Koch, reproduced in Table 2. 2.(146) 'Substantial' and 'almost perfect' , that is 0.61 – 1.00, were deemed to be acceptable.

Coefficient	Interpretation
< 0	Agreement worse than by chance
0.01 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost perfect

Table 2. 2 Landis and Koch (1977) suggested interpretation of agreement coefficients

2.2.4 Ordinal weighting

The laboratory and central code may disagree about the designation of a certain AKI stage or agree completely. AKI stages would be best described as being ordered i.e., AKI stage 2 is more severe than stage 1, and stage 3 is more severe than stage 1 or 2. There may, therefore, be partial agreement when the classification is adjacent (e.g., AKI is classified as stage 1 by the laboratory and stage 2 centrally) rather than non-adjacent (e.g., AKI is classified as stage 1 by the laboratory and stage 3 centrally). Weighting was used to penalize classifications that were further away from each other.(144) Ordinal weights were automatically chosen by the `kappaetc` command based on the number of potential classifications i.e., AKI stages 0 – 3. The weighting used is shown in Table 2. 3.

		Central coded AKI			
		0	1	2	3
Laboratory generated AKI	0	1.00	0.83	0.50	0.00
	1	0.83	1.00	0.83	0.50
	2	0.50	0.83	1.00	0.83
	3	0.00	0.50	0.83	1.00

Table 2. 3 Ordinal weights matrix applied to the primary analysis

Laboratory-generated and central-coded alerts that are in further disagreement from one another are weighted lower but not regarded as being totally in non-agreement, except at the extremes of staging, as staging is not binary but rather ordered by severity.

2.2.5 Exploratory analysis

Agreement was recalculated in several exploratory analyses to test various assumptions. These are outlined below.

Sensitivity analysis: Firstly, SCr values not associated with a laboratory alert were not recoded to AKI 0 as there might have been other reasons why an alert was not received (such as alert suppression or incomplete data submission). These laboratory alerts were, therefore, set to missing and were discarded from the complete-case analysis. Secondly, ordinal weights were not included so agreement was assumed to be binary (agreed/disagreed). Given that the algorithm was initialized at different times in different laboratories, and there may have been teething problems early on, agreement was also assessed over calendar time (yearly 2015 – 2020). Lastly, agreement was assessed for laboratories that submitted only complete data as there was concern that missing SCr data might conflate the observed results. Data were only designated complete if all months had non-missing alert and SCr data. The laboratory's data

was otherwise removed. Note that Green and Amber classified data, as detailed in section 1.4.5, only pertains to alert data; missing SCr are not currently classified by the UKRR so this definition of completeness was preferred instead.

Sub-group analysis: The algorithm would be expected to operate reliably in people with pre-existing CKD and at extremes of age. Checks were conducted to ascertain if the code was performing consistently amongst different quantiles of age and baseline SCr[9].

The algorithm is integrated into the laboratory information management system (LIMS) which is usually outsourced to a commercial provider. Agreement was evaluated across commercial LIMSs, when the LIMS provider was known. Theoretically, there are potentially two versions of code: firstly, that which is encoded by the LIMS provider and secondly, further modified by the laboratory. Agreement was thus examined for individual laboratories primarily.

2.2.6 Ethical considerations

The UKRR has ethical approval to use the surveillance data it receives for audit and research purposes without individual patient consent (under section 251 of the Health and Social Care Act, United Kingdom). Permission was granted by the UKRR Data Release Group within the UKRR for use of their data for this analysis based on a quality improvement initiative. Data were pseudonymised prior to data extraction. Patients were able to voluntarily withdraw their data, and this is processed internally within the UKRR.

[9] Baseline SCr is not available for laboratory alerts because they are not currently transmitted to the UKRR. Quantiles were calculated from the RV computed within the central algorithm code.

2.3 CONCLUSION

In this chapter, the methods used to investigate the consistency and implementation of the NHSE AKI detection algorithm were presented. AKI alerts generated from laboratories were compared with alerts generated by a simulated identical algorithm executed within longitudinal SCr data from persons with laboratory-submitted AKI alerts from before, during and after the AKI. Inter-rater agreement methods were used to quantify consistency of alerts and subgroup analyses were used to investigate consistency in different population groups.

In the following chapter, the results of this analysis and a discussion are presented.

3 CONSISTENCY OF ALERTS GENERATED BY, AND IMPLEMENTATION OF, THE NHS ENGLAND ACUTE KIDNEY INJURY DETECTION ALGORITHM IN ENGLISH LABORATORIES

3.1 INTRODUCTION

The validation of the NHS England algorithm and its implementation in English laboratories is imperative if alert data are to be used for AKI interventions and surveillance. This chapter compares local-laboratory alerts with alerts produced using code simulating the NHS England algorithm. Agreement was assessed using inter-rater methods, detailed in chapter two.

3.2 RESULTS

3.2.1 Baseline characteristics of the alerts included in this analysis

Out of 1,966,003 AKI alerts representing 532,884 patients received by the UKRR from the beginning of the NHS England algorithm deployment from December 2014 to the end of September 2020, 1,579,663 alerts – 475,634 individuals – were available for analysis; exclusions are shown in Figure 3. 1. Out of the 37 laboratories, eight were excluded. Two laboratories inconsistently sent data leading to frequent missing alerts and SCr values over time. There was evidence of missing pre-alert SCr data for five laboratories. One laboratory was excluded because only SCr data were sent but not any alerts. The median age was 72.4 years and 47% were female (Table 3. 1). The number of local laboratory alerts generated per person varied, averaging median 3 (IQR 1; 8) per person. AKI was frequently mild (63% were

AKI 1). The median baseline SCr in the year before the first laboratory alert was 77 µmol/L (IQR 57; 108). Laboratories included in this analysis were well represented across England (Figure 3. 2) although not all were included in this analysis as explained in section 1.4.5.

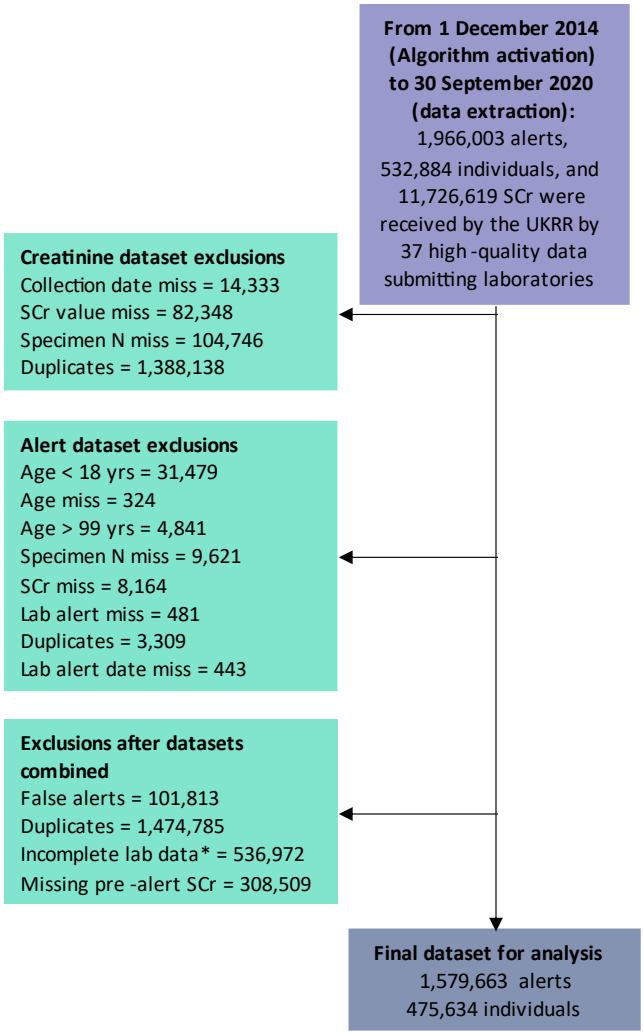


Figure 3. 1 Flowchart of participants and alert inclusion and exclusion in this analysis

Alert and SCr datasets are received separately by the UKRR. For the alert dataset, each line of data was an alert. The SCr dataset was provided as one line per SCr result. Exclusions therefore refer to the number of alerts or SCr results, as appropriate. Datasets were first cleaned and then re-cleaned after appending together. False alerts appeared to occur before a baseline SCr was available, indicating that some SCr values were not submitted or were otherwise unusable because of associated missing data. These were recoded to missing rather than deleted to avoid losing the associated SCr result. For duplicates between alert and SCr datasets, the duplicate occurring from the SCr dataset was dropped and the duplicate from the alert dataset was retained. Only the age < 18 years exclusion was pre-defined. Numbers do not exactly total as exclusions were dropped sequentially. *See section 3.2.1 for details. Abbreviations: **SCr**, serum creatinine; **lab**, laboratory; **miss**, missing; **N**, number.

	Local laboratory alerts, N = 1,579,663	**Central alerts,** N = 1,646,850
Age in y, median (IQR)	72.4 (59.6; 82.0)	
Sex, % Female	 47%	
Nonzero alerts The median number of alerts per person (IQR) Range AKI 1, N (%) AKI 2, N (%) AKI 3, N (%)	 3 (1; 8) 1; 252 988,284 (63%) 290,998 (18%) 300,351 (19%)	 4 (2; 9) 1; 352 1,067,239 (65%) 349,405 (21%) 230,206 (14%)
Median creatinine at alert (IQR) µmol/L All AKI AKI 1 AKI 2 AKI 3 Baseline SCr for the 1st alert* RV1 RV2 After the last alert	 164 (111; 264) 132 (96; 187) 181 (137; 237) 414 (307; 555) 77 (57; 108) - - 83 (60; 121)	 169 (111; 286) 140 (97; 219) 197 (139; 286) 361 (238; 577) - 83 (IQR 56; 132) 87 (IQR 64; 127) -
Before and after SCr results N before (the 1st alert) Median number per person N after (the last alert) Median number per person	 3,583,049 17 (9; 13) 2,025,767 13 (6; 26)	

Table 3. 1 Characteristics of alerts that were included in this analysis

*Computed for the first alert here, but baseline SCr values are not received by the UKRR. Abbreviations: **IQR**, interquartile range; **AKI**, acute kidney injury, **N**, number; **SCr**, serum creatinine; **RV1**, reference (baseline SCr) value for preceding 0 – 7 days; **RV2**, reference (baseline SCr) value for preceding 8 – 365 days.

Figure 3. 2 Map of England showing the laboratories that contributed to this analysis

Laboratory locations across England are shown. Map created using Leaflet (RStudio).

3.2.2 Characteristics of the centrally derived alerts

The central code computed a greater number of AKI nonzero alerts (1,646,850) than those received by the UKRR, indicating possible under-ascertainment or incomplete submission of alerts. Compared to local laboratory-generated alerts displayed in Table 3. 2, the SCr at central alert was higher than local laboratory-generated alerts for AKI 1 (median 140 vs 132 µmol/L) and AKI 2 (median 197 vs 181 µmol/L) but was 53 µmol/L lower for AKI 3 (median 361 vs 414 µmol/L).

The central code also enumerated the RV1[10], RV2[11] and C1: RV ratio[12], information that is not submitted to the UKRR for local laboratory-generated alerts so cannot be compared. The RV1, which could potentially be drawn during the prodromal illness before AKI is recognised, was lower than RV2 for AKI 0 (78 vs 86 µmol/L), but higher than RV2 for all other stages – AKI 1: 109 vs 90, AKI 2: 122 vs 89, and AKI 3: 161 vs 88 µmol/L. The C1: RV ratio, which is used to calculate relative changes in SCr, averaged 1.62 (IQR 1.53; 1.75) for AKI 1, 2.30 (IQR 2.12; 2.56) for AKI 2 and 3.81 (IQR 3.21; 5.13) for AKI 3. Although the baseline (RV1) was higher, the median C1: RV ratio for AKI 1 from between 0 – 7 days was < 1.5 suggesting that the relative change in SCr that flagged AKI 1 was from the 8 – 365-day period or an absolute increase in SCr of > 26 µmol/L in 48 hours.

[10] Reference SCr within 0 – 7 days of the index SCr.
[11] Reference SCr within 8 – 365 days of the index SCr.
[12] Index SCr to the highest of RV1 or RV2. The highest ratio is preferred.

3.2.3 Local laboratory versus central generated alerts

Table 3. 2 shows a cross-tabulation of local laboratory-generated alerts compared with those derived centrally. The local laboratory and central alerts mostly concurred. The majority were AKI 0.

	Centrally generated alerts				
	0	**1**	**2**	**3**	**Total**
Local lab generated alerts after the first ever alert					
0					
Percent	**79.81**	1.82	0.51	0.49	82.64
Frequency	**7,260,258**	165,802	46,541	44,433	7,517,034
1					
Percent	1.85	**8.67**	0.23	0.11	10.86
Frequency	167,922	**788,731**	21,288	10,343	988,284
2					
Percent	0.33	0.26	**2.52**	0.09	3.20
Frequency	30,252	23,711	**229,065**	7,970	290,998
3					
Percent	0.34	0.70	0.50	**1.77**	3.30
Frequency	30,522	63,430	45,053	**161,346**	300,351
Total					
Percent	82.33	11.45	3.76	2.46	**100.00**
Frequency	7,488,954	1,041,674	341,947	224,092	**9,096,667**

Table 3. 2 Matrix of raw numbers of laboratory versus centrally derived alerts

Laboratory 0 alerts were assumed on the basis that an alert was not generated for a given SCr value. This was only assumed for alerts generated after the first-ever alert for that laboratory was submitted as a proxy for the date of activation of the algorithm.

3.2.4 Inter-rater analysis of local laboratory versus central alerts

In general, agreement was almost perfect – overall Gwet's AC1 was 0.98 and 26 laboratories computed a Gwet's AC1 of >0.81 (Table 3. 3). Some alerts generated within laboratories showed only slight agreement with the central code; Gwet's AC1 ranged from 0.17 – 0.23 in three. This was driven by two patterns of misclassification in these three laboratories: (1) central AKI 0 was misclassified as laboratory AKI 3 in 38 – 55% of central AKI 0 alerts and (2) central AKI 0 was misclassified as laboratory AKI 1 in 88 – 92% of central AKI 0 alerts.

N lab alerts	N central alerts	Percent positive agreement	Gwet's AC1 agreement	95% CI	LIMS provider
18,912	5,589	0.6852	0.1697	0.1564, 0.1830	CSC
42,262	15,789	0.6992	0.1726	0.1638, 0.1814	CSC
35,318	9,869	0.7146	0.2261	0.2172, 0.2351	Unknown
9,126	5,438	0.9176	0.8518	0.8464, 0.8573	Unknown
118,433	91,156	0.9542	0.9241	0.9233, 0.9250	Unknown
81,974	65,234	0.9571	0.9428	0.9422, 0.9434	Unknown
54,069	68,744	0.9619	0.9460	0.9452, 0.9468	Unknown
21,522	29,385	0.9631	0.9497	0.9486, 0.9508	Unknown
45,720	31,098	0.9650	0.9521	0.9514, 0.9529	CliniSys
7,976	8,891	0.9737	0.9613	0.9595, 0.9631	CSC
87,505	104,631	0.9733	0.9631	0.9626, 0.9636	CliniSys
82,902	95,496	0.9772	0.9662	0.9657, 0.9667	CSC
42,813	51,575	0.9791	0.9677	0.9669, 0.9685	Unknown
84,134	84,948	0.9766	0.9682	0.9678, 0.9686	Unknown
25,552	24,729	0.9809	0.9702	0.9693, 0.9711	Unknown
4,408	8,232	0.9761	0.9702	0.9690, 0.9713	Unknown
75,392	102,496	0.9760	0.9703	0.9700, 0.9707	InterSystems
121,016	165,896	0.9806	0.9750	0.9747, 0.9753	Unknown
58,138	65,579	0.9826	0.9765	0.9761, 0.9769	CSC
37,312	41,825	0.9853	0.9793	0.9787, 0.9799	CSC
118,612	133,778	0.9852	0.9803	0.9800, 0.9805	CliniSys
26,784	30,183	0.9865	0.9808	0.9801, 0.9815	CSC
53,376	56,986	0.9902	0.9865	0.9862, 0.9868	CliniSys
110,469	118,988	0.9909	0.9866	0.9864, 0.9869	Unknown
26,617	27,624	0.9910	0.9879	0.9874, 0.9883	CliniSys
99,768	105,677	0.9916	0.9882	0.9880, 0.9884	Unknown
38,871	42,990	0.9938	0.9917	0.9914, 0.9919	Unknown
33,153	35,631	0.9943	0.9926	0.9923, 0.9928	Unknown
17,499	18,393	0.9959	0.9943	0.9939, 0.9946	Unknown

Table 3. 3 Agreement coefficients for anonymised individual laboratories

Sorted in ascending order of agreement. Coefficients are displayed to 4 decimal places as the CI was very narrow. Abbreviations: **lab**, laboratory; **LIMS**, laboratory information management system; **AC**, agreement coefficient; **CI**, confidence interval; **N**, number; **CliniSys**, Clinical Systems; **CSC**, Computer Science Corporation.

Laboratories often outsource their LIMS to a commercial provider which has taken the responsibility to incorporate the NHS England algorithm within its workflow. Table 3. 4 shows that assessment of agreement was also almost perfect (> 0.81) across all three LIMS providers; Gwet's AC1 ranged from 0.97 – 0.98. LIMS provider was only known for 43% of local laboratory alerts, however, since the provider was unknown for 16/29 laboratories. Of the three laboratories demonstrating slight agreement, two of these utilised the CSC LIMS provider, previously shown to be performing well at LIMS-level. This suggests a local laboratory rather than a LIMS inconsistency. Cumulatively, these represented a small proportion of the total alerts (8%).

	N laboratories	**N local laboratory alerts**	**Percent positive agreement**	**Gwet's AC1**	**95% CI**
CliniSys	4	331,830	0.9811	0.9745	0.9742, 0.9749
CSC	7	274,286	0.9659	0.9505	0.9499, 0.9511
InterSystems	1	75,392	0.9760	0.9704	0.9698, 0.9710

Table 3. 4 Agreement within different laboratory information management system providers

Only a limited number of laboratories are included here since the LIMS provider was not always available or the laboratory currently uses its own unique LIMS. Abbreviations: **CliniSys**, Clinical Systems; **CSC**, Computer Science Corporation; **AC1**, agreement coefficient; **CI**, confidence interval.

3.2.5 Exploratory analyses

Several exploratory analyses were conducted to examine particular aspects of the NHS England algorithm of clinical importance and practical application.

Firstly, given that there were laboratories that submitted incomplete alert data, these 11 laboratories were excluded. Table 3. 5 showed that by excluding laboratories with patchy missing data, agreement was no higher (Gwet's AC1 was 0.97).

	Centrally generated alerts				
	0	**1**	**2**	**3**	**Total**
Local lab generated alerts after the first ever alert					
0					
Percent	**79.48**	1.65	0.45	0.37	81.96
Frequency	**3,626,595**	75,268	20,638	17,108	3,739,609
1					
Percent	2.59	**8.37**	0.19	0.02	11.16
Frequency	118,393	**381,709**	8,554	730	509,386
2					
Percent	0.49	0.39	**2.45**	0.05	3.39
Frequency	22,504	18,000	**112,009**	2,218	154,731
3					
Percent	0.46	0.82	0.52	**1.68**	3.48
Frequency	21,195	37,240	23,847	**76,628**	158,910
Total					
Percent	83.04	11.23	3.62	2.12	**100.00**
Frequency	3,788,687	512,217	165,048	96,684	**4,562,636**

Table 3. 5 Matrix of numbers of lab versus central alerts for complete laboratories

Laboratories with missing alert or SCr data for any month were excluded.

AKI 3 is graded as more severe than AKI 2 and AKI 1. Ordinal weighting was therefore applied to the primary analysis although a binary totally agree/disagree evaluation may also be of interest. Without weighting alerts depending on how much they disagreed, agreement was lower, but still almost perfect, compared to the primary weighted analysis (Gwet's AC1 0.93 versus 0.97, compared in Table 3. 6).

	Percentage positive agreement	**Gwet's AC1**	**95% CI**
Ordinal weights applied	0.9756	0.9661	0.9660, 0.9762
Unweighted	0.9277	0.9196	0.9194, 0.9298

Table 3. 6 Comparison of agreement coefficients unweighted versus ordinal weighting

Ordinal weighting was applied to account for the natural ordering of AKI severity that increases with the KDIGO stage. Abbreviations: **AC**, agreement coefficient; **CI**, confidence interval; **KDIGO**, Kidney Diseases: Improving Global Outcomes.

The UKRR does not receive AKI 0 alerts and so local laboratory-generated alerts were not exactly comparable with alerts generated by the central code. For this reason, AKI 0 alerts were assumed on the basis that an alert was *not* received. Agreement, between nonzero AKI 1/2/3 alerts only, was once again almost perfect, though substantially lower, when local laboratory alerts were not recoded to AKI 0 – Gwet's AC1 was 0.83 versus 0.97. Alert agreement are compared in Table 3. 7 (number of alerts) and Table 3. 8 (agreement coefficients).

	Centrally generated alerts				
	0	1	2	3	Total
Local lab generated alerts					
1					
Percent	10.63	**49.93**	1.35	0.65	62.56
Frequency	167,922	**788,731**	21,288	10,343	988,284
2					
Percent	1.92	1.50	**14.50**	0.50	18.42
Frequency	30,252	23,711	**229,065**	7,970	290,998
3					
Percent	1.93	4.02	2.85	**10.21**	19.01
Frequency	30,522	63,430	45,053	**161,346**	300,351
Total					
Percent	14.48	55.45	18.70	11.37	**100.00**
Frequency	228,696	875,872	295,406	179,659	**1,579,633**

Table 3. 7 Matrix of the numbers of laboratory-generated versus centrally derived alerts excluding AKI 0 alerts

Zero alerts i.e., no AKI, had to be assumed if no AKI alert was received by the UKRR for that particular SCr value. Here, only AKI 1, 2 and 3 alerts generated by laboratories were compared to the central algorithm-generated alerts. Abbreviations: **lab**, laboratory.

	N alerts	Percent positive agreement	Gwet's AC1	95% CI
Local lab alerts included assuming missing alerts were AKI 0 (Local lab AKI 0,1,2, and 3)	9,096,667	0.9756	0.9661	0.9660, 0.9662
Local lab alerts not generated on the basis that they were not received by the UKRR excluded (agreement only between local lab AKI 1,2, and 3)	1,579,633	0.9197	0.8254	0.8247, 0.8262

Table 3. 8 Comparison of agreement coefficients for missing alerts versus assuming missing laboratory alerts were in fact AKI 0

Coefficients were weighted. Abbreviations: **CI**, confidence interval; **AKI**, acute kidney injury; lab, laboratory; **UKRR**, UK Renal Registry; **AC**, agreement coefficient; **N**, number.

The NHS England algorithm has not been altered since its implementation. However, it was hypothesised that agreement might not be the same over time as laboratories sequentially actively started submitting data to the UKRR and difficulties may have been encountered during the early stages of roll-out. An exploratory analysis, presented in Table 3. 9, found that Gwet's AC1, per calendar year, from 2015 to September 2020 was similar even during the 2020 Coronavirus-2019 pandemic (0.96 compared to 0.97 the previous year). SCr data submission was incomplete for seven laboratories in 2020 compared to one in 2019.

	N alerts	**Percent positive agreement**	**Gwet's AC1**	**95% CI**
2015	125,824	0.9681	0.9560	0.9556, 0.9564
2016	224,363	0.9734	0.9635	0.9632, 0.9638
2017	302,863	0.9761	0.9673	0.9671, 0.9676
2018	346,131	0.9784	0.9704	0.9702, 0.9707
2019	357,731	0.9781	0.9697	0.9694, 0.9699
2020	222,721	0.9726	0.9590	0.9586, 0.9593

Table 3. 9 Agreement over time, based on the year of laboratory-generated alert

The deadline for algorithm implementation was March 2015 and data extraction was up until September 2020. Not all laboratories activated timeously. Abbreviations: **AC**, agreement coefficient; **CI**, confidence interval; **N**, number.

It was expected that since CKD is common in the general population that the NHS England algorithm should perform consistently in people with and without CKD. Over quartiles of baseline SCr, agreement was high at lower SCr but decreased substantially at higher baseline SCr values. Gwet's AC1 was 0.98 for quartile 1 but dropped to 0.88 for quartile 4 as shown in Table 3. 10.

Quartile	Median SCr, µmol/L	Percent positive agreement	Gwet's AC1	95% CI
1	42	0.9820	0.9759	0.9757, 0.9761
2	65	0.9900	0.9861	0.9860, 0.9863
3	92	0.9832	0.9749	0.9747, 0.9751
4	164	0.9268	0.8765	0.8759, 0.8770

Table 3. 10 Agreement across quartiles of baseline SCr (as determined by the central algorithm)

Abbreviations: **SCr**, serum creatinine; **AC**, agreement coefficient; **CI**, confidence interval.

Moreover, for the highest baseline SCr, quartile 4, the frequency of AKI 1 was greatest for central compared to local laboratory-generated alerts (18% [central] versus 15% [local]) but lower for stage 3 alerts (4% [central] versus 9% [local]). This suggests that severe AKI might be under-reported in people with pre-existing CKD because alerts are suppressed by laboratories.

SCr is subject to non-GFR factors such as sarcopenia, haemodilution and certain medication use, which might be more prevalent in older people at risk of muscle loss, heart failure, and that use RAAS inhibitors and other medications. In the final exploratory analysis, agreement was found to be almost perfect (> 0.81) and consistent in younger and older people. In Table

3. 11, the median age for each quintile ranged from 45 – 88 years and Gwet's AC1 was between 0.81 – 0.85.

Quintile	Median age, years	Percent positive agreement	Gwet's AC1	95% CI
1 (18 – 55.9 years)	45	0.9223	0.8308	0.8290, 0.8322
2 (56 – 68.9)	63	0.9164	0.8112	0.8095, 0.8129
3 (69 – 76.9)	72	0.9151	0.8095	0.8078, 0.8112
4 (77 – 83.9)	80	0.9177	0.8230	0.8214, 0.8246
5 (84 – 99)	88	0.9269	0.8520	0.8506, 0.8534

Table 3. 11 Agreement across quintiles of age

Abbreviations: **CI**, confidence interval.

3.3 DISCUSSION

This study used routinely collected AKI alert and SCr data submitted to the UKRR to, for the first time, internally validate the NHS England AKI detection algorithm by assessing its implementation and consistency of AKI alerts generated by laboratories in England. Local-laboratory alerts submitted to the UKRR were compared to alerts produced by simulated code. There was almost perfect agreement between local-laboratory and centrally derived alerts although implementation of the algorithm may not have been consistent in some individual laboratories and when the baseline SCr was high.

3.3.1 AKI characteristics in the context of prior epidemiological research

Consistent with other studies of AKI in the UK, the patients represented in this analysis were of older age.(147) This analysis found that local-laboratory/ central algorithm agreement was comparable across quintiles of age which is reassuring since SCr can be affected by non-GFR factors which older people are especially at risk for.(148) The average baseline SCr was within the normal reference range (62 – 115 µmol/L for males and 45 – 84 µmol/L for females in most laboratories). This is most likely explained by a high prevalence of sarcopenia, especially given the average age, rather than a low prevalence of pre-existing CKD. In a previous UKRR analysis, up to 25% of people with AKI alerts had pre-existing CKD.(149)

A systematic review that investigated validation studies of administrative databases found that the sensitivity of coded-AKI compared to biochemical definition of AKI codes was low. Biochemical AKI was defined heterogeneously. AKI, previously called acute renal failure, had > 30 definitions. That review included studies that were conducted prior to the near universal and contemporary definition of AKI (KDIGO) and pre-dated modern information technology

systems.(150) More recently, validation of the NHS England algorithm compared alerts generated with clinical adjudication using RIFLE criteria(68) and ICD-10 coded-AKI.(67) The algorithm was sensitive (90.5% and 91.2%, respectively). Though, ICD-10 coded AKI tends to under-recognise AKI stage 1 which is itself associated with poor kidney outcomes and mortality.(39,151) AKI 1 was most frequently enumerated in this analysis. An advantage of using the NHS England automated algorithm is thus that it captures all stages of AKI uniformly.(70)

3.3.2 Potential clinical benefits of enumerating AKI and collecting AKI data by the UKRR

Although patient- and hospital-level randomised controlled trial evidence of interruptive AKI alerts alone or in combination with care bundles to improve quality of care and vital and kidney outcomes have been negative, there might be other clinical benefits.(56,152–154) Notwithstanding the use of alerts as real-time notifications, the utility of NHS England AKI detection for the purposes of surveillance is still valuable but demands generous resource allocation to collect and analyse these data.(63)

Registries collect data on kidney disease epidemiology and outcomes but are only common in high-income countries according to the Global Kidney Health Atlas survey, partly because of limited health information technology infrastructure.(155–157) The ISN has developed the SHARing Expertise to support the set-up of Renal Registries initiative (SharE-RR; see www.theisn.org). AKI registries are not practical unless AKI detection is automated. This study found that although the AKI data accumulated by the UKRR might be incomplete, the alerts are still valid given that, barring a few, laboratories were mostly generating alerts that agreed

with the central code. AKI alerts that the UKRR receives form a unique AKI registry which could potentially improve AKI clinical trial recruitment efficiency, diversity and follow-up.(152,158) Recently CKD and KRT clinical trial design has begun to exploit registries for patient recruitment and follow-up.(159) Evidence generated from these trials may have public health benefits depending on their results.

3.3.3 Good agreement within LIMS but poor agreement within some individual laboratories

The Think Kidneys best practice guideline (available at https://www.thinkkidneys.nhs.uk/aki/resources/clinical-biochemists/) illustrates the NHS England algorithm. Laboratories are mandated to follow this algorithm for the purpose of AKI detection, but since the syntax of any written code was not provided, the code integrated into LIMSs might not be identical. In this analysis, agreement was similar and almost perfect within three LIMS providers, although not all providers were known. This is encouraging since it indicates compliance with the prescribed guidance. It also suggests that any divergence in agreement found is likely due to individual laboratory implementation or failed data submission. Nevertheless, it should not be surprising that there might be misinterpretation of the algorithm by laboratories given that even nephrologists do not always agree about how KDIGO criteria should be clinically applied in database research.(160)

On the other hand, there was poor agreement within some individual laboratories. The three lowest performing laboratories demonstrated 'Green' data quality, defined by the UKRR, and implemented the algorithm in early 2016 and 2017. Nevertheless, these three laboratories over-enumerated AKI compared to the central code. This has implications for the UKRR which

reports AKI incidence rates. In the UKRR 2018 AKI report, the sex and age adjusted rate varied from 2,600 – 20,600 per million population.(72) The disagreement amongst a few laboratories in the present analysis might explain this variation although these three laboratories were not at the extreme incidences reported by the UKRR. Sawhney et al have previously demonstrated that when similar methodological rules and age-sex adjustments are applied, there are usually little true differences in regional AKI incidence in the UK.(147)

Some laboratories submitted incomplete data. Either there was missing information for certain variables or alert or SCr data were missing completely for one or more month(s). Exclusions used in this analysis attempted to eliminate the former and a secondary analysis of only complete laboratories was performed as an appraisal of the latter. Agreement was very similar for laboratories contributing complete monthly data, suggesting that the effect of missing monthly data is negligible. This observation is reassuring as missing data plagues research in general and in particular observational research that registry data may be used for. Others have experienced similar issues in missing data when detecting AKI in biochemistry and electronic health record data. To overcome some of these barriers, Sawhney et al for example amended their algorithm to use 2 days rather than the 48 hour rule to detect recent SCr changes in lieu of date-time.(67) Johnson et al used linked data across several databases to limit incongruous data.(161) The UKRR does attempt to reconcile demographic data with NHS Digital which might help to limit missing and remedy inaccurate records by cross-referencing and filling in missing information submitted to the UKRR with information held by NHS Digital. Although the NHS England algorithm is automated, the submission of alert and SCr data to the UKRR is manual so it is impossible to estimate the extent of missing data. It is

also difficult to recover, beyond sending regular reminders to laboratories. Automated data transmission processes may overcome this obstacle.

Think Kidneys permits suppression of alerts in certain circumstances such as alerts originating from neonatal and renal units. Also, their best practice guideline notes that: *"The effect of previous AKI episodes on the diagnosis of new AKI episodes"* and *"Increases of serum creatinine over 48 hours of more than 26µmol/L (but less than 50% above baseline) in stable chronic kidney disease (CKD)"* should be addressed to reduce false positive alerts.(72) No specific recommendation is offered by the guideline allowing different interpretations of how to address these concerns. LIMSs and laboratories might adopt potentially incongruent code modifications. Alternatively, the code may be similar, but alerts following the first alert, at which AKI is detected for the first time, may be suppressed and not sent to the UKRR. This might explain why agreement was very low for some and high for other laboratories which should be employing equivalent code and serviced by the same LIMS provider. 'Exception' reports should be made compulsory and forwarded to the UKRR so that they may be scrutinised. Rules should be standardised in the Think Kidneys guideline to ensure laboratories are implementing comparable code.

3.3.4 Misclassification of AKI as worsening CKD

In this analysis, it was found that agreement fell at high baseline SCr values. There were twice as many AKI 3 alerts generated by local-laboratories compared to the central code – since there were more rather than less alerts computed, laboratory suppression of alerts does not explain this finding. This indicates that the NHS England algorithm might be flagging

worsening CKD as AKI. Previous external validation of the algorithm found that up to 14% of CKD was misclassified as AKI (when compared to nephrologist adjudication).(68)

False positive AKI has previously been found to be more common in people with higher baseline SCr (>1.5mg/dL [133 µmol/L]) particularly because a 26.5 µmol/L (0.3mg/dL) absolute increase, as included in defining KDIGO stage 1 AKI, is easier to reach at high SCr values when biological and analytical variation in SCr is higher.(70,162). The balance between detecting true cases and disregarding false positive AKI is delicate; criteria can be modified to increase sensitivity at the expense of specificity.(67,163) For example, Sawhney et al extended the look-back period of the NHS England algorithm to 3 years, which increased the sensitivity of the algorithm but also increased the number of CKD that was flagged as AKI.(67) While not practical for notifying clinicians in real-time, sensitivity also increased when AKI was detected retrospectively. For example, the nadir after the peak SCr might be used to indicate that AKI occurred. This could be useful when there is an unknown baseline SCr and is still practical for submission to the UKRR.(67,70) Nevertheless, future iterations of the NHS England algorithm might require redesign for use in people with CKD and there should be clarification of how alerts generated by people with CKD are processed within LIMSs.

Moreover, the UKRR does compare lists of people starting long-term dialysis with the AKI database to remove possible false positive alerts generated by changes before/ after maintenance dialysis, but this information is only received periodically unlike the AKI data that is received monthly causing a delay in removal of false alerts. Unfortunately, acute dialysis is not captured by the NHS England algorithm, as recommended by KDIGO to stage AKI 3.(11)

A related theoretical misclassification problem was recognised which will be investigated in a future analysis. The NHS England algorithm compares a median SCr up to 365 days (RV2) in addition to the lowest SCr within 7 days (RV1).(11) This potentially inflates the baseline SCr to a higher value when preceding high SCr results associated with previous AKI are used to calculate the baseline that triggers future AKI alerts. Using a 'fixed' baseline as investigated by Lin et al, might generate fewer misclassified alerts. (162)

3.3.5 Strengths and limitations

This study captured a substantial number of alerts and individuals over an extended period. There were, however, some limitations as well. The dataset available for this analysis was a subset of high-quality relatively complete data from laboratories regularly sending files to the UKRR. This analysis should be repeated on all 190 laboratories. Even so, those that were included were geographically representative of all laboratories in England. Also, this study used routinely collected data that notoriously are subjected to issues of missing data and duplicate entries. The UKRR, however, performs extensive data cleaning, reconciliation, and clarification with NHS Digital and communicates with laboratories directly to resolve irregularities.

Moreover, the NHS England algorithm does generate AKI 0 alerts, but these, as well as the baseline SCr and threshold used to generate alerts, are not currently submitted to the UKRR. These metadata would be invaluable to make further comparisons that were not possible in this present analysis.

Lastly, there may be bias in how SCr is assayed – the difference in SCr results may differ by 10% between Jaffe and enzymatic methods.(164,165) A survey of UK laboratories in 2022 found that 31% of respondents reported that the modified kinetic Jaffe was still being used.(166) The UKKA eGFR Working Group has since issued a PSA stating all UK laboratories should transition to enzymatic methods. Since laboratories serve the same GP practices and hospitals, at least there would be consistency in the SCr assay and so relative changes are likely to be valid. However, caution is advised once laboratories replace the assay with enzymatic methods, historical results, to which the current SCr is compared to, might trigger AKI because of analytical variability. Unfortunately, the dataset used in this current analysis did not have information on the assay used for each individual laboratory.

3.4 CONCLUSION

The NHS England algorithm remains a reliable tool to detect AKI. Alert generation locally, rather than centrally at the UKRR, is acceptable. The utility of the AKI data that the UKRR collects as part of the AKI registry is vital for quality improvement projects and monitoring AKI burden across England. In addition, UK Renal Registry-based trials of AKI can be assured that AKI alerts are valid. However, it would be helpful if the UKRR received more complete and additional information from laboratories. Further scrutiny is required of a few laboratories and in people with pre-existing CKD.

In the next chapter, the clinical epidemiology of AKI is investigated in the City of Cape Town, using a bespoke AKI detection algorithm used by the Provincial Health Data Centre (of the Western Cape). The NHSE algorithm, investigated in this chapter, is also compared to the bespoke algorithm.

4 CLINICAL EPIDEMIOLOGY OF ACUTE KIDNEY INJURY IN THE CITY OF CAPE TOWN, SOUTH AFRICA, USING ROUTINELY COLLECTED ADMINISTRATIVE DATA

4.1 INTRODUCTION

The currently available literature on the burden, risk factors and outcomes of AKI in South Africa are limited to single-centre studies and highly selected hospital-based populations such as people living with HIV, pregnancy and sepsis-associated AKI, following trauma or major non-cardiac surgery, patients admitted to critical care and requiring acute dialysis, and AKI that is referred for nephrologist evaluation which is often more severe.(42,59,167–172)

The objective of this study was to describe the clinical epidemiology of AKI in the City of Cape Town across all disciplines and settings, using routinely collected data from primary, secondary, and tertiary levels of care collated by the PHDC, with AKI episodes identified using a universally applied AKI algorithm. A secondary objective was to compare the AKI detected by the PHDC AKI algorithm with AKI alerts identified by the NHSE detection algorithm, applied in chapter three, and a modified version of that algorithm which will be described later in section 6.2.3.

This chapter is divided into three sections: section 4.2 gives the methods used for the analysis, section 4.3 details the results of the analysis and in section 4.4 the results are discussed in the

context of the general population, contemporaneous literature and the limitations of the algorithm used to detect AKI and available data.

4.2 METHODS

4.2.1 Study Design and Definitions

This study was a descriptive prospective virtual cohort study. The cohort is described as 'virtual' as no direct contact was made with the participants because the data were routinely collected administrative data from the healthcare service.

Study population: Data from healthcare clients with newly PHDC-ascertained AKI aged > 18 years old during a 5-year period (1 January 2017 to 31 December 2021) were extracted for all public healthcare encounters to primary, district, regional and tertiary levels of care in the City of Cape Town, South Africa. This prolonged period was chosen to record temporal trends and ensure the capture of kidney disease and SCr results that would otherwise be missed if encounters with the healthcare system were infrequent. A summary of the PHDC was given in section 1.4.10 and the AKI algorithm was described in section 1.4.11. Suspicious dates of birth that may have been captured incorrectly, for example, where an individual's age at the time the AKI episode was > 100 years or year-of-birth 1900, were excluded. Only the year of birth was provided in the dataset, therefore the day and month were imputed to the 1st of July as the midpoint for the purpose of the age calculation.

Measurement of serum creatinine: The National Health Laboratory Service (www.nhls.ac.za) provides all chemical pathology testing to the public healthcare sector in South Africa. These

laboratories use the modified kinetic Jaffe method traceable to isotope dilution mass spectrometry to measure SCr using various platforms (personal communication with Dr Jody Rusch, chemical pathologist). The limit of detection is 5 – 2,700 µmol/L but values outside of this range are diluted out as per protocol (version 13.0 Cobas® package insert). Laboratories are located on-premises at all district-, regional- and tertiary-level hospitals in the City. Primary healthcare clinics, intermediate care, psychiatric and tuberculosis facilities send samples for testing to the nearest laboratory. SCr values < 5 or > 4000 µmol/L were excluded.

Definition of an acute kidney injury episode: AKI was ascertained using the PHDC's detection algorithm, introduced in section 1.4.11. Commencement of the episode occurred on the date on which criteria were first met for AKI and ended on the day criteria were no longer met, as per the PHDC algorithm. This means that an AKI episode was allowed to potentially accrue for many weeks or months until an SCr reached < 100 µmol/L. AKI episodes remained open if no subsequent SCr < 100 µmol/L was reached. When known, open episodes were closed using the in-facility date of death.

AKI that developed in the community (community-acquired AKI, CA-AKI) was compared to hospital-acquired AKI (HA-AKI). HA-AKI was defined as occurring after the first two days of admission, that is, on day three, up until the last date of the admission. This was chosen instead of 48 hours as the time stamp of the admission and AKI episode was unknown. All other AKI was characterised as CA-AKI.

Definition of comorbidities and relationship with the episode: Diabetes, hypertension, and human immunodeficiency virus (HIV) infection are prevalent conditions in South Africa and

the Western Cape and were ascertained by the PHDC using specific evidence of such conditions, outlined in section 1.4.10. (79,173–175) Comorbidities were deemed present if they occurred before or during the AKI episode. Importantly, ascertained comorbidities only indicate that a comorbidity had been captured using existing digital records, an uncaptured condition was not evidence of its absence. Hypertension is ascertained using the dispensing of specific drug treatments and so untreated hypertension and whether hypertension is controlled was also unknown.

4.2.2 Outcomes

Characteristics of people who developed AKI included demographics (age at the time of the first AKI episode and sex[13]) and comorbidities. Other attributes associated with the episode were described, such as episode duration, average SCr values, frequency of SCr testing, facility type (primary healthcare clinic and district, regional and central hospital) and visit encounter type (primary healthcare visit, hospital outpatient and admission) before and on the same day as the AKI episode ascertainment. In addition, for individuals with CA-AKI who were hospitalised and all individuals with HA-AKI, the duration of hospital admission length of stay was described.

Healthcare utilisation: Encounters with the healthcare system are captured using the facility administrative systems. Visits to primary healthcare facilities, district, regional and tertiary hospital outpatient appointments, emergency centre and admissions were described. Emergency visits were evidenced by a speciality code of “emergency medicine” and the visit

[13] Birthdate and sex are captured by facility clerks during patient registration using the national identification document or passport, if available, otherwise it is self-reported by the health-seeker or their chaperone.

type “HECTIS” (Hospital and Emergency Centre Tracking Information System). Not all AKI would have been captured as ‘emergencies’ since not all emergency visits are seen via emergency medicine or in the emergency centre. For example, gynaecology and obstetrics emergencies usually are seen in a dedicated unit and referrals may be admitted via specialities directly. Nephrology services are provided by the two national central hospitals in the Western Cape, Groote Schuur Hospital and Tygerberg Hospital.

Mortality: Vital status and date of death were determined using the hospital administrative system (Clinicom™). The provincial Department of Health cannot access the national death register except in some situations with linkage via the national ID number, but the national ID is only recorded in ~30% of PHDC registrants, and in addition many undocumented migrants who do not have South African IDs make use of public healthcare facilities.(176) As such, only in-facility death could be described with confidence for hospitalised CA-AKI and all HA-AKI. Cause of death data are not captured in the PHDC data, and thus the mortality described in this study represents all-cause mortality. Death was plotted as the proportion that had evidence of death during the same month of an admission as the AKI episode as only the month and year of death were available. Since only in-facility death was known, encounter data was used to infer that the patient was at least *not* deceased because of evidence of facility attendance after the start of the AKI episode – this was described after 30-days and 1-year after the start of the AKI episode.

4.2.3 Analysis

Demographic, creatinine, AKI characteristics, comorbidities and outcomes were described using appropriate summary statistics of measures of frequency (counts and proportions) and central tendency and dispersion. CA- and HA-AKI episodes were compared using descriptive statistics. Stata version 17 (StataCorp, TX, USA) was used for data cleaning, description, and visualisation.

Cross-boundary migration between South African provinces, the non-attendance at healthcare facilities of healthy individuals, and the influence of private healthcare sector beneficiaries, who are not captured by the PHDC, or who move back and forth between the public and private healthcare systems precludes an accurate calculation of the denominator population. As such, only the absolute number of AKI events could be described using these data, and not the incidence rate or prevalence in the general population.

Exploratory analysis: The AKI ascertained by the PHDC using their bespoke algorithm was compared with AKI detected using the NHSE algorithm, and a modified version of the NHSE algorithm developed to detect acute-on-chronic kidney disease. The latter will be described in detail in section 6.2.3. A summary of the algorithm attributes is given in Table 4. 1.

	Algorithm		
	PHDC	**NHSE**	**Modified NHSE**
Baseline creatinine	Most recent result or retrospectively re-assigned if no result pre-AKI available	Updates with each subsequent new SCr measurement Not imputed if missing No retrospective assignment	Fixed for the duration of the AKI episode Not imputed if missing No retrospective assignment
Criteria for AKI detection	SCr ≥ 100 µmol/L	KDIGO	KDIGO (11)
Recovery assessment	SCr < 100 µmol/L: infinite assessment window	None	Per Sawhney et al (177)

Table 4. 1 Comparison of amended algorithm with the NHSE and PHDC AKI detection algorithms

Abbreviations: **PHDC**, Provincial Health Data Centre (of the Western Cape); **NHSE**, National Health Services England; **SCr**, serum creatinine; **AKI**, acute kidney injury; **KDIGO**, Kidney Diseases: Improving Global Outcomes.

The algorithms were executed in the SCr data associated with all PHDC ascertained AKI (1 January 2017 to 31 December 2021); see Figure 4. 1. Comparisons were made between the algorithms using inter-rater agreement methods, as in section 2.2.3.(144) Because the PHDC does not generate alerts for AKI as such, a dummy variable for each SCr result was created and coded to one on the start of AKI detected by the PHDC (AKI_{PHDC}) until the end of the episode, and null otherwise. Because AKI_{PHDC} does not elucidate the AKI stage, but the NHSE (AKI_{NHSE}) and modified NHSE ($AKI_{modified\ NHSE}$) algorithms do, a dummy variable one/ null was similarly created for SCr associated with AKI_{NHSE} alerts and a one/ null dummy variable was

created for $AKI_{modified\ NHSE}$ alerts. These dummy variables were then compared across each SCr result and agreement assessed.

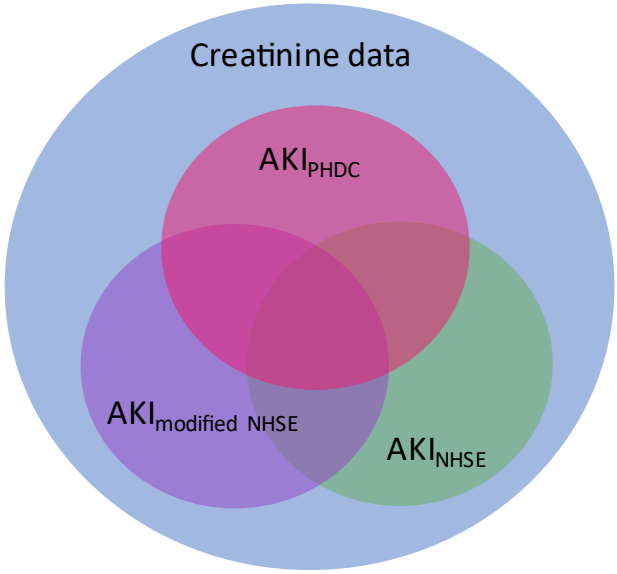


Figure 4. 1 Illustration of comparison of AKI alerts and data used to derive the alerts

Serum creatinine data from healthcare clients with AKI_{PHDC} 365 days prior to the individual's first AKI_{PHDC} alert until the end of 2021, including the SCr results during the AKI_{PHDC}, were used to compare alerts generated by the PHDC with AKI_{NHSE} and $AKI_{modified\ NHSE}$ algorithms. Overlapping areas signify agreement between algorithms whereas non-overlapping areas signify alerts that were exclusively generated by that algorithm but not by another. Abbreviations: **AKI**, acute kidney injury; **PHDC**, Provincial Health Data Centre (of the Western Cape); **NHSE**, National Health Services England.

4.2.4 Ethical considerations

Ethical clearance was obtained from the University of Cape Town Human Research Ethics Committee (2020/765) for the use of PHDC data for patients with SCr data from the City of Cape Town. Before data were released by the PHDC, data were de-identified and dates of birth and death were made incomplete to ensure patients are not traceable or re-identifiable. A waiver of individual patient consent was granted by the Ethics Committee because of the anonymised and perturbed nature of the data provided.

4.3 RESULTS

4.3.1 PHDC ascertained AKI

During the study period from 1 January 2017 – 31 December 2021, after excluding n=4,182 (age < 18 years) and n=33 (age > 100 years) individual healthcare clients, 158,965 AKI episodes in adults were identified by the PHDC; 135,730 unique healthcare clients with unique identifiers were represented. The annual number of episodes over time was similar 2017 – 2020 except in 2021 when this increased by 25% compared to 2020 (Table 4. 2). Most individuals only had a single documented AKI episode (1 episode: 74.3%, 2: 17.1%, 3: 5.2%, 4+: 3.4%). For ease of interpretation, only the first episode will be described hereon.

Characteristics of the episodes are available in Table 4. 2. For the first-ever episode, the median age was 50.1 years (IQR 37.2, 63.6) and 33.2% were of female sex. The proportion of females to males increased as the age increased: 32% female in age category 18 – 39 years, 31% in 40 – 64 years, 37% in 65 – 74 years and 49% in ≥75 years. Only 25% had no pre-existing comorbidity, whereas 44% had ≥ two known comorbidities. Episodes were prolonged > 7 days in 82.2%.

	PHDC AKI episodes, n episodes = 158,965	**Community-acquired,** n episodes = 138,122	**Hospital-acquired,** n episodes = 20,843
N patients with at least one episode	135,730	123,328	17,820
Age category, %			
18 – 39 years	30.4%	30.4%	30.8%
40 – 64	47.1%	47.2%	45.7%
65 – 74	14.2%	14.1%	15.3%
≥75	8.3%	8.3%	8.2%
Female sex@, n (%)	33.2%	31.6%	48.3%
Number of episodes, n (%)			
Per year			
2017	31,331 (19.7%)	27,429 (19.9%)	3,902 (18.8%)
2018	29,784 (18.7%)	25,713 (18.6%)	4,066 (19.5%)
2019	30,813 (19.4%)	26,425 (19.1%)	4,388 (21.1%)
2020	28,672 (18.0%)	24,310 (17.6%)	4,362 (21.0%)
2021	38,365 (24.1%)	34,280 (24.8%)	4,085 (19.6%)
Per quarter year			
Jan-Mar	44,068 (27.7%)	38,884 (28.1%)	5,184 (24.9%)
Apr-Jun	38,496 (24.2%)	33,386 (24.2%)	5,110 (24.6%)
Jul-Sep	37,730 (23.7%)	32,303 (23.4%)	5,427 (26.1%)
Oct-Dec	38,671 (24.3%)	33,589 (24.3%)	5,082 (24.4%)
Creatinine, median (IQR) µmol/L*			
Pre-AKI **^**			
2 days	84 (70, 96)	87 (73, 97)	83 (67, 95)
7 days	80 (65, 93)	84 (71, 96)	78 (63, 91)
365 days	74 (60, 88)	77 (63, 89)	70 (56, 84)
At ascertainment	154 (116, 293)	151 (115, 294)	166 (123, 289)
Average during episode	121 (103, 198)	121 (103, 203)	126 (103, 194)
Peak during episode	164 (117, 355)	161 (116, 356)	184 (128, 345)
Nadir during episode	100 (83, 118)	101 (84, 119)	94 (78, 114)
Post-AKI§	103 (75, 172)	105 (77, 180)	94 (69, 142)

SCr testing patterns			
Frequency in 365 days before AKI, median (IQR)	6 (2, 42)	21 (3, 85)	2 (1, 7)
Frequency in days during AKI, median (IQR)	3 (1, 43)	4 (1, 62)	2 (1, 4)
Frequency in 365 days after AKI, median (IQR)	3 (1, 24)	3 (1, 33)	2 (1, 7)
N SCr tests (%)			
2017	207,786 (21.1%)	69,275 (21.6%)	138,511 (20.9%)
2018	199,780 (20.3%)	69,299 (21.6%)	130,481 (19.7%)
2019	211,535 (21.5%)	71,511 (22.3%)	140,024 (21.1%)
2020	179,710 (18.3%)	60,766 (19.0%)	118,944 (18.0%)
2021	184,017 (18.7%)	49,718 (15.5%)	134,299 (20.3%)
Comorbidities*, yes %			
Hypertension	40.9%	41.0%	40.3%
HIV	30.9%	31.7%	24.1%
Diabetes	23.2%	23.0%	25.0%

Table 4. 2 Characteristics of people with AKI ascertained by the PHDC

@unknown for 0.09% of healthcare clients. *at the time of the first-ever episode. ^ 5,633 healthcare clients with CA-AKI had at least one SCr result in the preceding 7-days, 42,580 had at least one SCr result in the preceding 8 – 365-days; 8,936 healthcare clients with HA-AKI had at least one SCr result in the preceding 7-days and 8,483 had at least one SCr result in the preceding 8 – 365-days. § average SCr over the 365-days post-AKI end. Abbreviations: **AKI**, acute kidney injury; **CKD**, chronic kidney disease; **IQR**, interquartile range; **KDIGO**, Kidney Diseases: Improving Global Outcomes; **HIV**, Human Immunodeficiency Virus.

The PHDC does not specifically record the baseline SCr for each identified AKI episode, but the average value pre-AKI was lowest for the 365 days preceding the episode and highest in the 2 days prior to the episode. There were 10,648/154,958 (6.7%) individuals who later developed CKD with a median time of 39 days (IQR 1, 202) from first AKI to CKD ascertainment. The short interval between AKI and CKD ascertainment was likely explained by the fact that retrospective removal of the AKI ascertainment did not occur following confirmation of CKD at the second consecutive eGFR < 60 ml/min/1.73m^2 (CKD start date was at the first eGFR < 60 ml/min/1.73m^2). This is supported by 2,714 AKI episodes with an AKI start date after the CKD start, despite AKI not being permitted by the AKI_{PHDC} algorithm once CKD had been ascertained.

Creatinine characteristics of the first episode are shown in Table 4. 2. In addition, the median index SCr (at time of AKI ascertainment) was higher for females compared to males (127 [IQR 109, 183] versus 113 [105, 141] µmol/L). The average time to a subsequent AKI episode was 142 days (IQR 26, 529) and 1.4% of episodes recurred within 7 days of the end of the first AKI episode.

Compared to the UK Renal Registry 2020 report (data collected in 2018), healthcare clients with PHDC-ascertained AKI were younger, consistent with the younger general population of South Africa, and more frequently male, but episodes were similarly distributed throughout the year (Table 4. 3).(72)

	PHDC ascertained AKI in 2018, n episodes = 29,784; n patients = 27,229	**UKRR AKI alerts reported in 2018, n episodes** = 564,738; n patients = 488,856
Age category		
<18 years	-	2.3%
18 – 39	31.1%	8.7%
40 – 64	46.1%	21.8%
65 – 74	14.6%	20.0%
≥ 75	8.2%	47.2%
Female sex, %	32.5%	52.2%
N episodes by quarter		
Jan-Mar	27.6%	26.2%
Apr-Jun	24.4%	24.2%
Jul-Sep	22.6%	23.9%
Oct-Dec	25.3%	25.6%

Table 4. 3 Comparison with UK Renal Registry 2020 AKI report

Demographics and number (N) of episodes by quarter enumerated by the PHDC (City of Cape Town) compared to the AKI episodes reported by the UK Renal Registry 2020 report (2018 data). Abbreviations: **AKI**, acute kidney injury; **PHDC**, Provincial Health Data Centre; **UKRR**, United Kingdom Renal Registry.

4.3.2 Community- versus hospital acquired AKI

Of all PHDC ascertained AKI, 20,843/ 158,965 (13.1%) were hospital-acquired i.e., occurred after the first two days of an admission. This equates to 17,734 patients with at least one HA-AKI. CA-AKI accounted for 138,122/ 158,965 (86.9%) of all AKI, of whom 71,004 (51.4%) were subsequently hospitalised within 7 days.

Pre-AKI SCr in the preceding 2, 7 and 365 days was consistently lower for HA-AKI (Table 4. 2). AKI detection is wholly dependent on SCr testing (in the absence of other information on urine output and acute dialysis). Only 4.6% and 34.8% of first-ever CA-AKI episodes had a pre-existing SCr result in the preceding 7 days and 8 – 365 days, respectively. For HA-AKI, 66.8%

and 63.4% had a 7 day and 8 – 365 day preceding SCr result, respectively. Of those with pre-AKI results, a single SCr was recorded in the preceding 7 days in 72.9% for CA-AKI and 30.2% for HA-AKI. In the preceding 8 – 365 days, of those with pre-AKI values, a single SCr was recorded in 28.8% for CA-AKI and 9.0% for HA-AKI. The frequency of pre-AKI SCr testing was every 20 days (IQR 3, 84) for CA-AKI and 2 days (IQR 1, 7) for HA-AKI.

Severity: The index and peak SCr during the hospital-acquired episode were on average higher compared to CA-AKI (Table 4. 2). In addition, during the episode, regularity of testing was every 4 days (IQR 1, 62) for CA-AKI and 2 days (IQR 1, 4) for HA-AKI.

The AKI_{PHDC} algorithm closes the episode once SCr < 100 µmol/L. There was no SCr < 100 µmol/L by the end of data extraction in 59.0% of CA-AKI and 46.2% of HA-AKI. The median duration of the AKI episode was 20 days (IQR 3 – 146) in those with CA-AKI and median 4 days (IQR 2, 17) for HA-AKI.

Comorbidities: co-existing hypertension, HIV and diabetes mellitus were common at the time of AKI (Table 4. 2). More individuals with CA-AKI had HIV than HA-AKI, but other comorbidities were distributed similarly between community- and hospital-acquired AKI.

Encounters with the healthcare system before and during ascertainment: Before the CA-AKI episode, most healthcare clients did not encounter the healthcare system prior to their episode; of those that did, visits to primary healthcare facilities were the most frequent (Figure

4. 2). In contrast, most had an encounter prior to their HA-AKI; the majority of encounters were district-level hospital outpatient visits (Figure 4. 3).

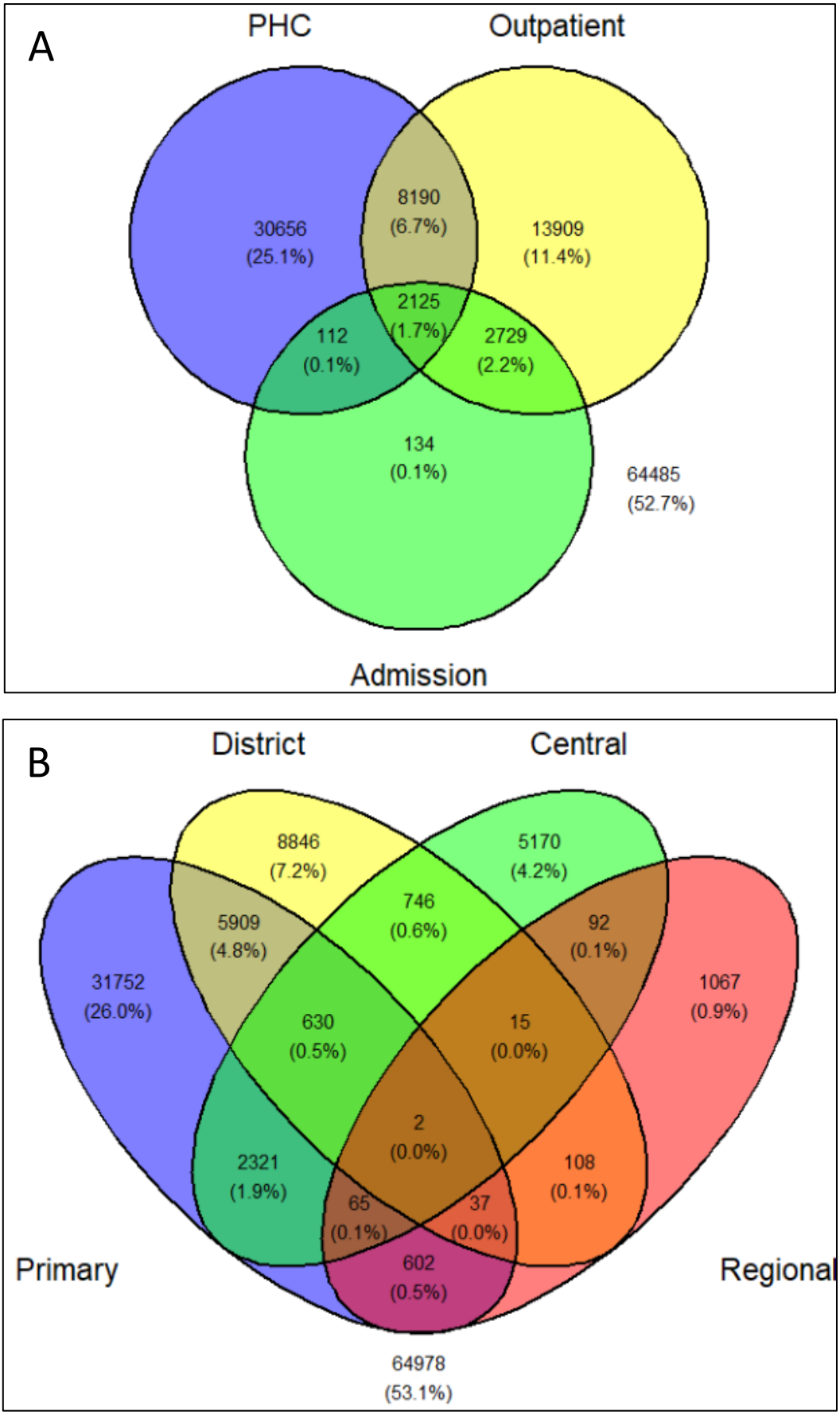


Figure 4. 2 Healthcare visit engagement and facility type prior to the CA-AKI episode

Intersecting circles are shown for (A) the number of primary healthcare (PHC), hospital outpatient and hospital admission visits and (B) primary healthcare, district hospital, regional hospital and central hospital facilities healthcare clients engaged with in the 30 days prior to their index CA-AKI episode. The number of the remaining individuals with no evidence of any healthcare engagement prior to their index CA-AKI episode are shown outside of the Venn. Nineteen individuals had an 'other' facility encounter but are not shown due to the number of intersecting ellipses possible. Other facility types included maternity obstetric units, oral health, psychiatric and tuberculosis hospitals, intermediate care, and satellite clinics. Drawn using the R ggvenn package. Abbreviations: **CA-AKI**, community-acquired acute kidney injury; **PHC**, primary healthcare.

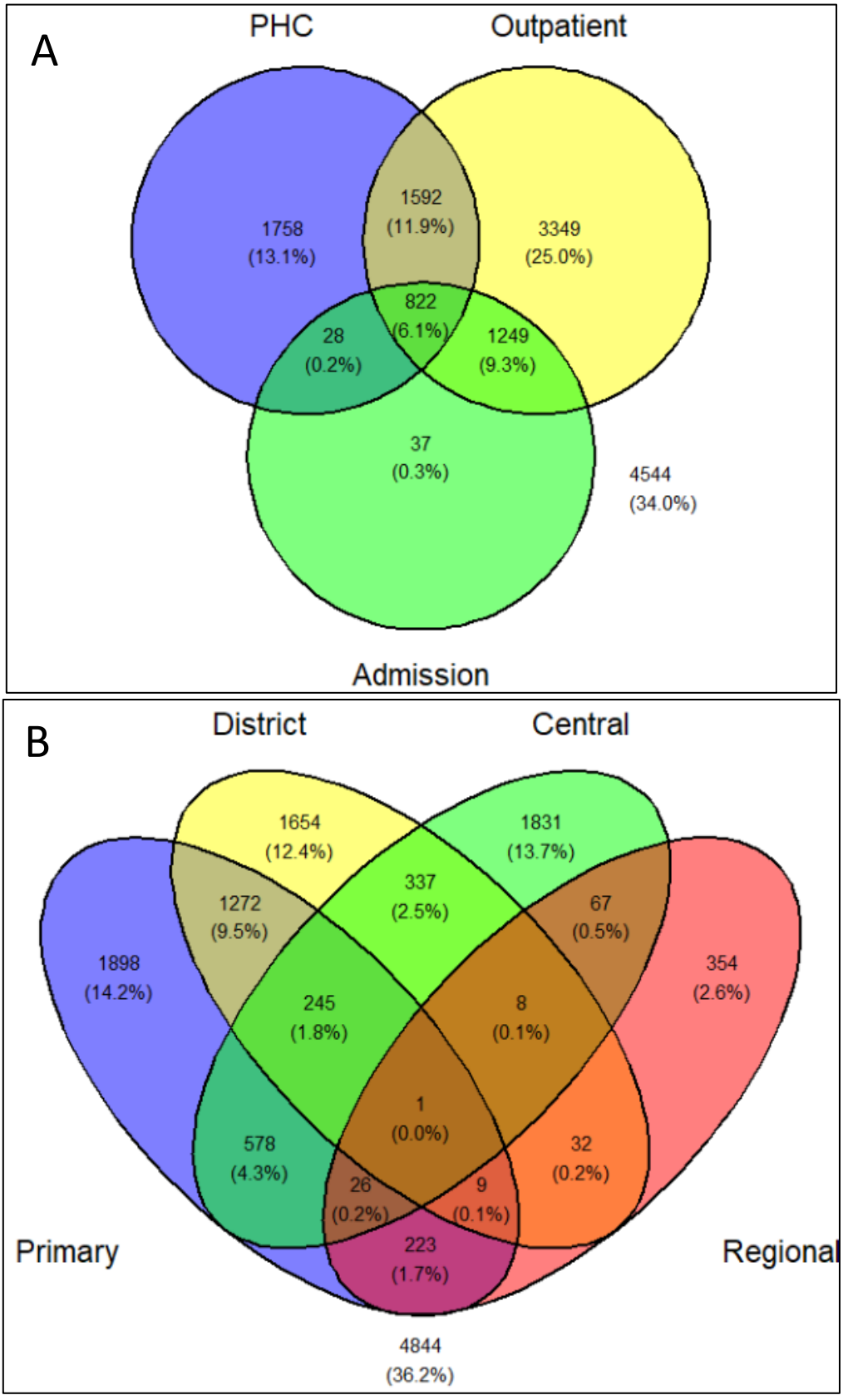


Figure 4. 3 Healthcare visit engagement and facility type prior to the HA-AKI episode

Intersecting circles are shown for (A) the number of primary healthcare (PHC), hospital outpatient and hospital admission visits and (B) primary healthcare, district hospital, regional hospital and central hospital facilities healthcare clients engaged with in the 30 days prior to their index HA-AKI episode. The index admission associated with HA-AKI was disregarded as the time prior to the AKI was of interest and admission was necessary for the definition of HA-AKI. The number of the remaining individuals with no evidence of any healthcare engagement prior to their index HA-AKI episode are shown outside of the diagrams. Nine individuals had an 'other' facility encounter but are not shown due to the number of intersecting ellipses possible. Other facility types included maternity obstetric units, oral health, psychiatric and tuberculosis hospitals, intermediate care, and satellite clinics. Drawn using the R ggvenn package. Abbreviations: **CA-AKI**, community-acquired acute kidney injury; **PHC**, primary healthcare.

In addition, CA-AKI was mostly ascertained at the time of a hospital visit (district: 39.7%, regional: 4.8%, central hospital: 24.4%, primary healthcare: 30.3%). HA-AKI was ascertained most frequently at central hospitals (62.3%) rather than other secondary levels of care (district: 26.1%, regional hospital: 4.1%)[14]. At CA-AKI ascertainment, 48.1% of hospital outpatient visits were emergency centre visits.

Length of stay for admissions to hospital that were associated with an AKI episode was median 4 days (IQR 2, 9) for CA-AKI ascertained within 7 days of the admission and 15 days (IQR 8, 25) for HA-AKI.

Kidney and vital outcomes: post-AKI, the average SCr was lower for HA-AKI than CA-AKI (Table 4. 2). The time to CKD ascertainment after the first AKI episode was only slightly longer for CA-AKI vs HA-AKI (community: median 39 days (IQR 1, 203), hospital: median 33 days (IQR 4, 193).

The proportion of individuals who had evidence of survival after 30 days after the start of the CA-AKI episode was 31.4% and after 1 year, 10.2%. For HA-AKI, however, 14.7% had evidence of being alive at 30 days but only 2.5% were alive at 1 year. Of those with hospitalised CA-AKI, 19.8% died during the same admission, mortality was highest for increasing age (Figure 4. 4) and the mean age at in-facility death was 55.5 (SD 17.0) years. Just over a quarter (28.0%) of HA-AKI had an in-facility death (Figure 4. 4); the average age at in-facility death was 53.0 (SD 16.1) years. Vital outcomes were unknown for the remaining group who had no evidence of in-facility death or no evidence of future healthcare engagement.

[14] The remainder occurred during long-term intermediate care facility admissions.

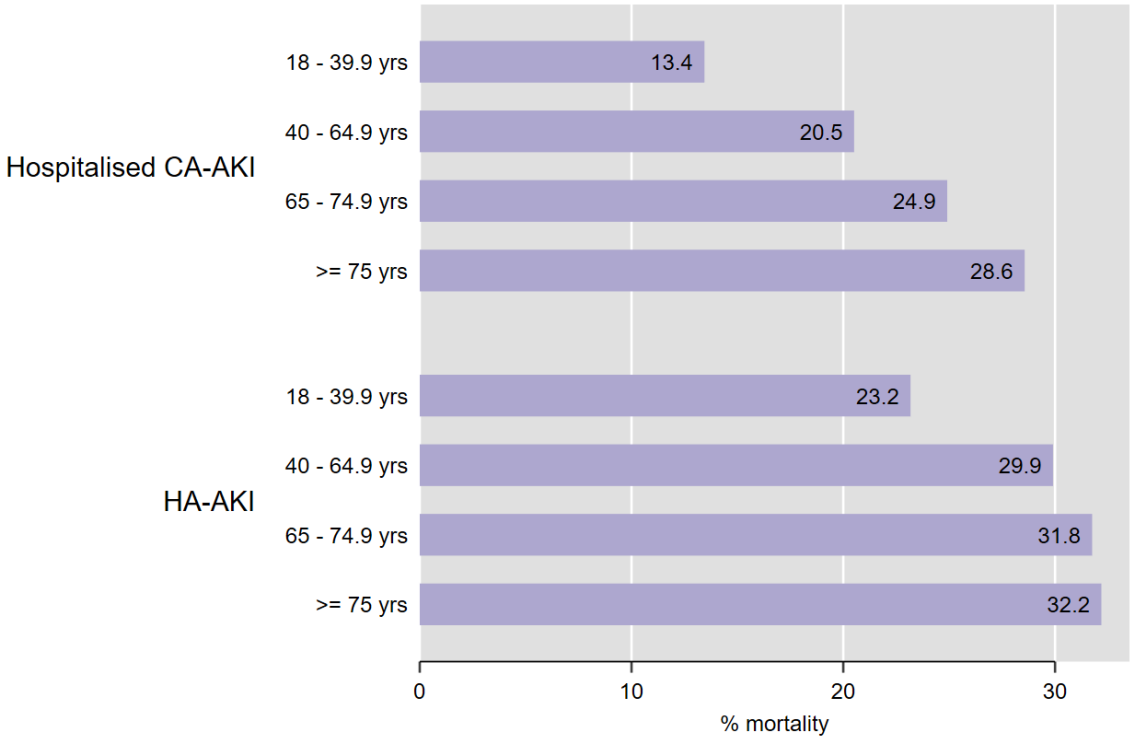


Figure 4. 4 In-facility mortality following community- and hospital-acquired AKI

The percentage of healthcare clients with hospitalised CA- and HA-AKI who demised during the hospital admission are shown by age category. In-facility death increased exponentially for CA-AKI, whereas the proportion who died with HA-AKI was similar except for the youngest age category. Only hospitalised CA-AKI mortality is shown since only in-facility death data were available. Abbreviations: **AKI**, acute kidney injury; **CA-**, community-acquired; **HA-**, hospital-acquired.

4.3.3 Comparison of NHSE, PHDC and amended algorithms

AKI alerts were dummy-coded from the duration of the PHDC ascertained AKI episodes, such that if there was an available SCr during the AKI period, an alert was artificially generated. Of 731,310 total SCr results for those with PHDC-ascertained AKI, 356,046 AKI_{PHDC} alerts were created.

In comparison, in the same creatinine dataset, the AKI_{NHSE} algorithm enumerated 119,732 AKI alerts and the $AKI_{modified\ NHSE}$ generated 95,235 alerts. There was perfect agreement between the AKI_{NHSE} and $AKI_{modified\ NHSE}$ algorithms, Gwet's AC1 was 0.9417 (95% CI 0.9411, 0.9424). However, agreement was only fair when the AKI_{PHDC} algorithm was compared to the AKI_{NHSE} algorithm (Gwet's AC1 0.3496; 95% CI 0.3473, 0.3519) and $AKI_{modified\ NHSE}$ (Gwet's AC1 0.3127; 95% CI 0.3103, 0.3151). Stage agreement was perfect between the AKI_{NHSE} and $AKI_{modified\ NHSE}$ (ordinal weighted Gwet's AC1 0.9723; 95% CI 0.9713, 0.9732) explained in part by the fact that the baseline creatinine values used to generate alerts were mostly comparable (Figure 4. 5). The staging alerts were compared in Table 4. 4.

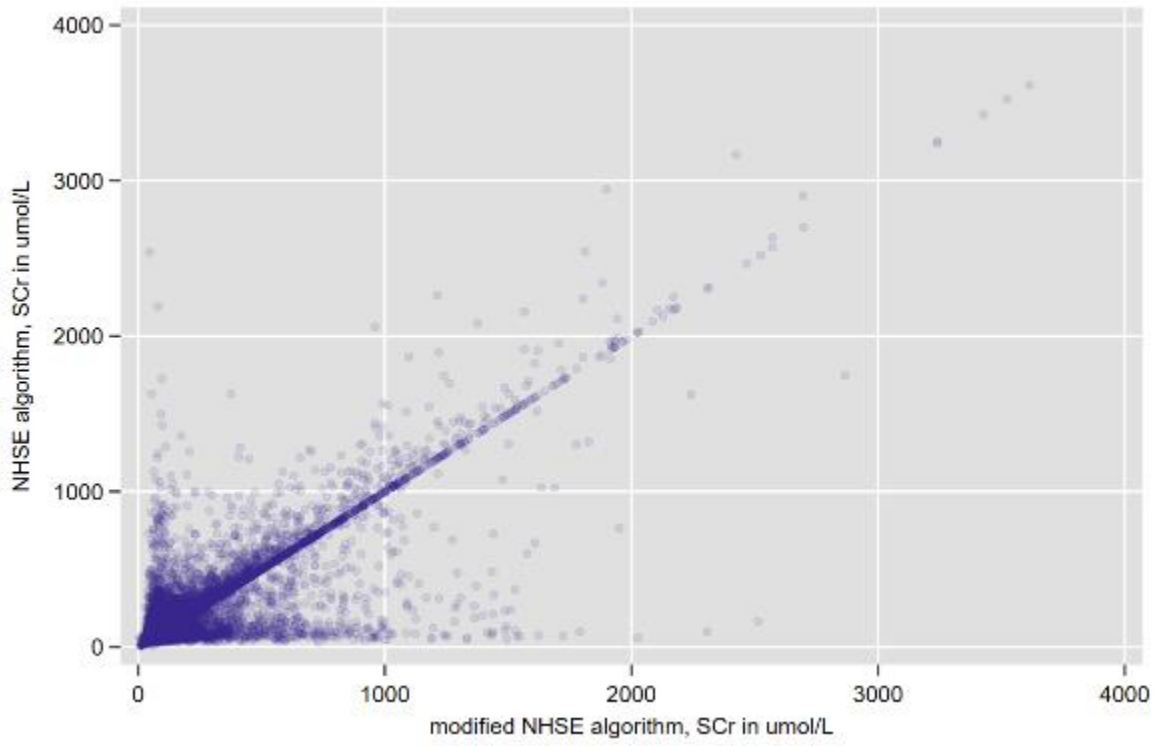


Figure 4. 5 Comparison of baseline creatinine used by NHSE and modified NHSE algorithms

Scatterplot comparing the baseline creatinine used to trigger AKI using the NHSE algorithm, which updates the baseline with successive SCr values during the AKI episode, and the modified NHSE algorithm that fixes the baseline to pre-AKI values for the duration of the AKI episode and resets the baseline assessment period after the AKI episode has closed (7 days after the start of the episode). Despite being updated with SCr values from during the episode, the baseline SCr values that triggered AKI were very similar. Only a random sample of 10% of all points is shown to remedy overplotting. High SCr values are caused by some individuals with CKD being included. Pearson's correlation = 0.85, $P < 0.001$. Abbreviations: **AKI**, acute kidney injury; **NHSE**, National Health Services England.

	Modified NHSE algorithm				
	0	1	2	3	Total
NHSE algorithm					
0					
Frequency	**608,336**	2,519	452	271	611,578
Percent	**83.18**	0.34	0.06	0.04	83.63
1					
Frequency	14,084	**35,864**	1,059	133	51,140
Percent	1.93	**4.90**	0.14	0.02	6.99
2					
Frequency	6,025	789	**22,807**	576	30,197
Percent	0.82	0.11	**3.12**	0.08	4.13
3					
Frequency	8,630	929	1,073	**27,763**	38,395
Percent	1.18	0.13	0.15	**3.80**	5.25
Total					
Frequency	637,075	40,101	25,391	28,743	731,310
Percent	87.11	5.48	3.47	3.93	100.00

Table 4. 4 Assessment of AKI algorithms comparing KDIGO staging alerts between NHSE and modified NHSE algorithms

In purple, the diagonal shows the majority of AKI alerts generated by both the NHSE and modified NHSE algorithms were of the same KDIGO stage. '0' alerts represent AKI criteria that were not met. Abbreviations: **NHSE**, National Health Services England; **AKI**, acute kidney injury.

4.4 DISCUSSION

For the first time, this study used routinely collected data on a virtual cohort of healthcare clients to describe the clinical epidemiology of AKI in public healthcare clients in the City of Cape Town, South Africa from 2017 to 2021, inclusive. Laboratory, facility encounter, and inferred comorbidity data collated by the PHDC health information exchange were used to characterise AKI that is detected using a pragmatic algorithm. The majority of AKI was community-acquired, which was not as severe as HA-AKI. Individuals were young and mostly male but multimorbidity was common. The young age of people experiencing AKI is in contrast to high-income countries, but is consistent with South Africa's younger population and also the demographic of people who experience AKI more frequently caused by infectious diseases, toxins, obstetric complications and trauma in Africa.(41,42,50,167,178)

4.4.1 The City of Cape Town in Context

As introduced in section 1.4.8, the City of Cape Town is the most populous district in the Western Cape Province. Province-wide, the PHDC has on record approximately 8 million active public healthcare clients with unique patient master index identifiers in the past decade whereas provincial population estimates for the Western Cape was 6.3 million in 2016.(79,178) A possible explanation is the high migratory behaviour of people moving between the urban City and more rural areas in neighbouring provinces, and the existence of many high-density informal settlements which are unlikely to be accurately enumerated in formal population estimates. These individuals are very likely to be reliant on the public health service, and are probably more likely to experience a higher burden of disease requiring public health service access than those in more affluent areas who are also more likely to access

private health care exclusively. The latest population estimate for the City of Cape Town – the largest provincial district and the focus of this analysis – was 4.6 million in 2020.(179)

4.4.2 Attributes of acute kidney injury episodes

CA-AKI was more numerous compared to HA-AKI, as is consistent with other parts of the world and Africa.(42,163,180) HA-AKI was, however, more severe than CA-AKI – using the peak SCr and higher mortality risk as proxies of severity. The length of hospital stay for HA-AKI was correspondingly longer compared to hospitalised CA-AKI.

The baseline SCr in closer proximity to the episode was higher than the average in the preceding year, suggesting that the injurious process had already commenced in the days preceding AKI ascertainment. This observation is important in the context of HA-AKI versus CA-AKI as CA-AKI predominantly tends to be triggered by changes in SCr that are measured from 8 – 365 days before the AKI compared to HA-AKI which is often triggered by changes within 48 hours and 7 days, as demonstrated in a Scottish regional population study.(163)

Although it was not possible to determine the incidence of CA-AKI versus HA-AKI, a previous analysis conducted at Tygerberg Hospital in the City of Cape Town, using the NHS England algorithm, found that amongst patients admitted to that hospital in a 6-month period, 6.2% of admissions were associated with HA-AKI.(59)

In that study of HA-AKI in a single central hospital only, 59.1% were female, about 10% more than found in the present study across district, regional and central hospitals. The 100 µmol/L threshold, as used by the PHDC to define AKI, may operate differently in females as the

reference range for SCr is lower, 49 – 90 µmol/L for females whereas it is 64 – 104 µmol/L for males (National Health Laboratory Service reference range). PHDC ascertained AKI may therefore capture more severe AKI in females than males since in some cases the SCr must more than double to trigger an AKI_{PHDC} episode. Indeed, the index SCr was found to be higher for females compared to males. In addition, males with a normal SCr above 100 µmol/L may be identified as AKI. Metanalytic evidence would suggest female sex is protective against AKI (OR for male sex: 1.23; 95% CI 1.11, 1.36) and especially so for HA-AKI (OR for male sex: 1.52; 95% CI 1.26, 1.70).(44) This would explain the low proportion of female sex overall, but the current observation of a near 1:1 male to female ratio for HA-AKI is possibly explained by hospitalised complications of pregnancy and the way women access healthcare differently to men. Kister et al. back-calculated the SCr from the CKD-EPI eGFR equation in females assuming male sex (the female coefficient was removed). Their results suggested that females with AKI are under-detected. (181) It would have been important, had data on non-AKI patients been available as well, to assess how a sex-specific threshold would change the case detection of AKI using the upper limit of the laboratory reference ranges – 90 µmol/l for women and 104 µmol/l for men – given the apparent disparity of how the 100 µmol/l threshold may operate differently in males and females.

Furthermore, the SCr > 100 µmol/l is problematic because, in the South African context, SCr may be affected by non-GFR factors most relevant to LLMICs. Poor socioeconomic circumstances, which affect people accessing public healthcare, make diets that contain animal-protein unaffordable, especially in vulnerable groups.(182,183) Wasting syndromes are also common for e.g. as a complication of HIV infection and active tuberculosis, as well as

the growing epidemic of sarcopenic obesity.(184,185) Drugs used in the management of HIV and its sequalae, namely trimethoprim for the prophylaxis of pneumocystis pneumonia and dolutegravir, increase the SCr in the absence of kidney injury.(186,187) This may cause false classification of both AKI and CKD in PLHIV and poorer communities.

Hypertension and diabetes were as common as previously described in a prospective cohort study of referred AKI conducted at Groote Schuur Hospital, and a population health study of urbanised black people in Cape Town highlighting the strong threat of diabetes and hypertension as risks for AKI.(42,188,189) Although blood pressure and glycaemic control were unknown in the present analysis, 116,726 healthcare clients with diabetes were recorded by the PHDC in 2015 – 2020 using a bespoke diabetes detection algorithm in the province and glycaemic control was relatively poor (57% had a glycated haemoglobin > 8%).(190)

The estimated population HIV burden in males aged 15+ in 2017 – 2021 increased from 6.5% – 6.8% and 11.3 – 12.3% in females in the City.(191) Thus, HIV as a comorbidity prior to AKI was substantially higher than the general population highlighting HIV, its treatment and complications, being strong risk factors for AKI and poor outcomes.(45) Dlamini et al found HIV to complicate 20.6% of hospitalised AKI at Groot Schuur Hospital, which was lower than the proportion with co-existing HIV in the current study, possibly explained by non-referral of PLHIV to nephrology services because of a perceived belief of poorer outcomes or other cognitive biases.(42) Another explanation worth consideration is that PLHIV undergo regular guideline directed kidney function testing and are therefore likely to be detected with AKI earlier and more frequently than HIV negative individuals.(192)

4.4.3 Healthcare utilisation and outcomes

The proportion of different types of healthcare encounters and frequency of SCr testing should be interpreted similarly: the more contact with the healthcare system for routine care, the more likely AKI will be detected (at least biochemically). In a Veterans Affairs study of AKI in the community in the USA, the hazard of CA-AKI identification was higher for more frequent outpatient utilisation (HR 2.38; 95% CI 2.31, 2.46).(193)

This is consistent with the many primary-level healthcare visits observed prior to AKI episodes. However, the fact that the majority of AKI was ascertained at hospital visits, especially as emergencies, highlights that AKI detection is not purely detected incidentally during routine visits.

Dedicated post-AKI clinics do not exist in South Africa, and the nephrology service is only likely to follow up AKI requiring acute dialysis. In a Canadian centre, a post-AKI follow-up clinic reduced the risk of death, although not CKD progression.(194) This is not practical in the setting of limited resources and scarce nephrologist workforce as is the case in South Africa.(195) Survival after 30-days and 1-year, proxied by healthcare engagement, was much lower and documented mortality was likewise higher for HA-AKI. While follow-up appointments may not have been scheduled, especially for the sole purpose of kidney function re-testing, someone may have disengaged with the healthcare system because of loss to follow-up, migration or transition to private healthcare, which may all be directly related to poor kidney health. Only in-facility death confirmation was obtainable by the PHDC and so mortality may well be underestimated.

4.4.4 The COVID-19 pandemic

The study period included the COVID-19 pandemic (March 2020 – ongoing) which may have affected key results. The number of episodes enumerated in 2020 was similar to historic values. This is despite SARS-CoV-2 being associated with multiorgan dysfunction. AKI complicated 28.6% (95% CI 19.8, 39.5) of COVID-19 related hospital admissions in an early world-wide meta-analysis.(196) In the year 2021, a surge of CA-AKI episodes was seen in this analysis, higher than previous pre-pandemic years and 2020, but little change in HA-AKI.

There was therefore likely an underestimation of AKI ascertainment in 2020 but also reduced inciting events such as interpersonal and road-traffic violence, especially since alcohol sales were prohibited and movement was hindered, and elective surgeries were cancelled.(197) Periods of public health and social measures otherwise known as “lockdowns”, in which movement of people was curtailed to help prevent the spread of SARS-CoV-2 may also have affected transport and access to facilities which would have had manifold implications for access to routine visits, missed diagnosis of AKI and hindered ability to monitor for AKI recovery, as was seen for HIV care.(198) At Mitchell’s Plain District Hospital, City of Cape Town, emergency centre visits decreased by 19% in 2020 compared to previous years.(199) The cause of AKI may also have shifted as there was disruption of elective surgeries and a reduction in trauma related injuries.(200,201) Also, creatinine testing requests from antiretroviral clinics reduced by 64% in 2020 compared to 2019 although 2018 to 2019 saw a reduction of 70% as well.(202) Although this was likely for routine monitoring of otherwise well individuals, kidney disease was potentially missed as a result of decreased testing in the current analysis.

4.4.5 Strengths and limitations

This analysis used the PHDC to characterise AKI in healthcare clients accessing public healthcare in the City of Cape Town. This is notable for two reasons. First, the PHDC is unique to the Western Cape; there is no analogous health informatic database that exists outside of the province in the rest of South Africa or Africa. Second, the laboratory data the PHDC receives allows the implementation of an AKI algorithm which is a highly efficient mechanism to detect AKI. Together, this allowed, for the first time, a rich epidemiological description of AKI attributes in the City of Cape Town. By using algorithms instead of classic manual audit, episodes are much more consistently ascertained but as highlighted, outcomes are difficult to confirm when healthcare clients disengage with care or do not receive dedicated AKI follow-up. Using routine health data in this way provides a large-scale view of the epidemiology of AKI in the whole population of public healthcare clients in the City of Cape Town, although it cannot provide the accuracy and deep phenotype that would be provided through a clinical study with recruited participants.

Although pragmatic, the rules used to detect AKI were not KDIGO aligned.(11) It was noted that the AKI_{PHDC} algorithm was not comparable to the AKI_{NHSE} and $AKI_{modified\ NHSE}$ algorithms. Further work will be undertaken in the future to provide evidence to the PHDC to assist with strengthening the AKI_{PHDC} algorithm, ensure that it is standardised to other validated algorithms and permit recognition of AKI in people with pre-existing non-KRT CKD. Despite theoretical concerns about the risk of a rolling baseline assessment period, the detection and severity staging of AKI was comparable between algorithms with and without a fixed baseline assessment period.

The PHDC only receives data from public healthcare facilities, and although many City residents access this sector or move between the public and private healthcare sectors according to their needs and financial means, exclusive private healthcare beneficiaries could not be included in this study to remain compliant with the Protection of Personal Information (POPI) Act of South Africa. This is also problematic for the current analysis because healthcare clients may periodically utilise private healthcare services to supplement their healthcare needs. Thus, data about individuals with AKI as well as their important encounters and comorbidities recognised in the private setting will not be acquired by the PHDC. In the future, it is envisioned that health insurance data will be onboarded within the PHDC, once appropriate consent structures can be implemented.(79)

Furthermore, because these data reflect the sector of the population actively seeking health care, the dataset is biased by the over-representation of clients coming for HIV, contraceptive and maternal wellbeing visits in the population represented in the PHDC data. This bias is also reflected by a paucity of health visits by healthy men. People with AKI are often otherwise ill and will present to facilities irrespective of routine visits so this bias is likely to be small.

4.5 CONCLUSION

This analysis of AKI in the City of Cape Town using routinely collected health data from public healthcare clients confirmed the high volume of both community- and hospital-acquired AKI episodes experienced and highlights the burden on hospital services, even in CA-AKI. Individuals who developed AKI were younger than the profile of patients with AKI in the United Kingdom. Despite being young, they had substantial multi-morbidity. These results should be interpreted in the context of the current AKI algorithm employed by the PHDC and the demographic of healthcare clients accessing public healthcare in the City. Further work is needed fine-tune the algorithm that is currently used by the PHDC to detect AKI but it is nonetheless promising that such an algorithm can be embedded in systems using laboratory data in South Africa.

In the next chapter, the clinical epidemiology of CKD in the City of Cape Town will be described.

5 CHARACTERISTICS OF HEALTHCARE CLIENTS WITH CHRONIC KIDNEY DISEASE IN THE CITY OF CAPE TOWN, SOUTH AFRICA

5.1 INTRODUCTION

Studies of CKD in South Africa have mostly thus far been limited by definitions using a single recorded eGFR and have been conducted in highly selected populations.(203,204) Consequently, using routinely collected data from healthcare clients accessing public healthcare services in the City of Cape Town, the characteristics of people with CKD were described in this analysis. This chapter is divided into methods, results and discussion sub-sections.

5.2 METHODS

5.2.1 Study design

This study was a prospective cohort study. The structure and sources of routinely collected data warehoused by the PHDC were previously introduced in section 1.4.10.

5.2.2 Study population

CKD episodes ascertained by the PHDC in adults aged > 18 years old during the period 2010 to 2021 in the City of Cape Town were included. The laboratory information management system was transitioned to the current software around 2010 and so eGFR data, necessary for the definition of CKD, were unavailable to the PHDC prior to this time in the current format.

The day and month of birth for all individuals was set to the 1st July as only the birth-year was provided by the PHDC for this analysis to prevent patient re-identification.

5.2.3 Definition of CKD

The PHDC defines CKD as two consecutive eGFR values < 60 ml/min/1.73m^2, at least 90 days apart, and was used as definitive evidence of CKD. The ascertainment date was the date of the first result of eGFR < 60 ml/min/1.73m^2. Albuminuria data are not currently available for inclusion in the definition or staging of CKD.(11) Evidence of KRT was used as supportive confirmation for CKD, as described in section 1.4.10. The episode was left open for a patient's lifetime, and closed only with the date of death, if known. CKD was staged at the time of ascertainment using KDIGO eGFR criteria as follows: eGFR <15, 15.0 – 29.9, 30.0 – 44.9, 45 – 59.9 ml/min/1.73m^2.(26) The NHLS provides the PHDC with eGFR values by MDRD equation without ethnicity correction.(4)

5.2.4 Definition of comorbidities

The proportion of healthcare clients with hypertension, diabetes and HIV were described. Definitions of comorbidity ascertainment using PHDC data were described in section 4.2.1.

5.2.5 Analysis

It was not possible to determine the population incidence or prevalence rate for CKD for three reasons, relating to the difficulty of estimating the total population as a denominator to calculate these rates. Firstly, the PHDC receives data from public healthcare facilities only and excludes private healthcare beneficiaries. Also, those who are physically healthy and are not visiting health facilities accordingly are under-represented by the PHDC. Secondly, kidney function testing is usually requested by indication either as directed CKD screening in people

at risk of developing CKD (such as individuals with diabetes and/or hypertension) or who are suspected of having kidney disease (based on the presenting complaint or bedside investigations). National HIV guidelines also recommend periodic kidney function testing in healthy people living with HIV (PLHIV) on antiretroviral therapy (to detect tenofovir nephrotoxicity) and in ill PLHIV who are at risk of complications of potential kidney disease due to AIDS-defining illnesses and their treatments.(205) The routine kidney function testing performed in South African public healthcare is thus very different to other settings in which incidence or prevalence may usually be described using routine visits.(192) Thirdly, there is significant internal migration between the City of Cape Town and the rest of the province and neighbouring provinces. Also, it is suspected that individuals seek healthcare from different sub-districts. For example, exemplary HIV and tuberculosis care have been established in the Khayelitsha sub-district and non-Khayelitsha residents may choose to access these services there. An attempt to estimate the rates of disease would therefore not provide a fair representation of the general population. Instead, incidence and prevalence trends over time were described as counts within the context of the healthcare-seeking population rather than the general population.

Individuals with newly ascertained CKD per calendar year 2010 – 2021 were counted as incident. Period prevalent CKD was counted as the number of healthcare clients in that calendar year that were newly or previously ascertained with CKD and had no evidence of death in that same year. Mortality captured by the PHDC are mostly in-facility deaths that are recorded in the facility administration system. Deaths out-of-facility or out-of-province will therefore be missed. The prevalent population was therefore also described separately by using evidence of survivorship. Evidence of a future encounter with the (public and provincial)

healthcare system – facility visits, medication dispensing and laboratory testing – was used as a proxy of survivorship. Individuals who transferred care to the private sector or emigrated out of the province would still be uncaptured. The two ways in which the prevalent population was derived are illustrated in Figure 5. 1.

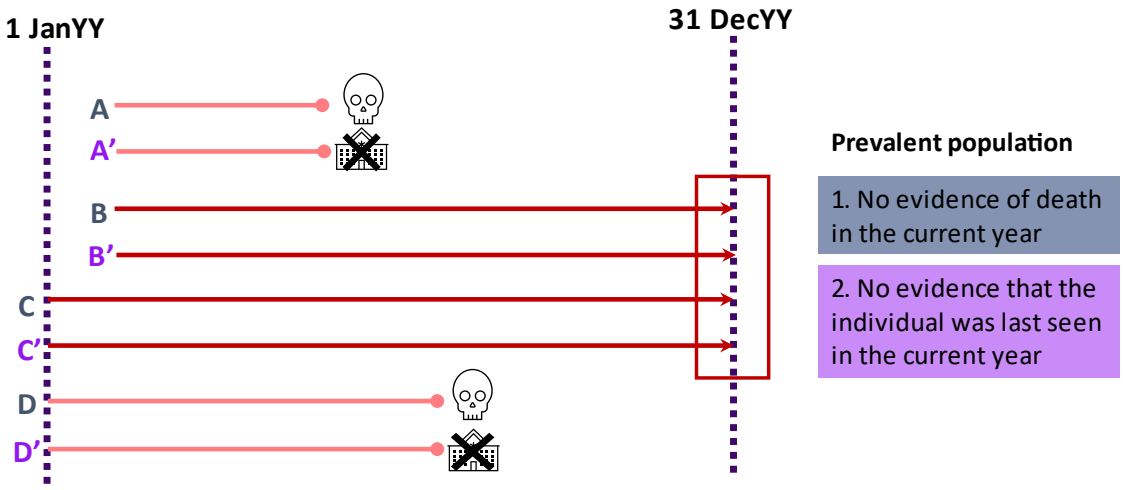


Figure 5. 1 Illustration of the derivation of the prevalent CKD population

A: newly ascertained in that year and died, **A'**: newly ascertained in that year and last seen in that year, **B**: newly ascertained in that year and no evidence of death, **B'**: newly ascertained in that year and evidence of healthcare engagement after the index year, **C**: previously ascertained with CKD and no evidence of death in that year or earlier, **C'**: previously ascertained with CKD and evidence of future healthcare engagement after the index year, **D**: previously ascertained with CKD and affirmative evidence of death in that year, **D'**: previously ascertained with CKD and no evidence of future healthcare engagement after the index year. Timelines ending in a filled circle were not analysed whereas arrow-headed groups were used to derive two prevalent populations, as shown in the boxes. Abbreviation: **CKD**, chronic kidney disease.

Summary statistics were used to describe the demographics, evidence of KRT and comorbidities of individuals with ascertained CKD. Only the 2019 period prevalent population was described in detail due to the observation that there were substantially lower numbers of newly ascertained CKD cases in 2020 and 2021 and limited time accrued before data

extraction to definitively determine that no future encounter was experienced to confirm individuals that were no longer prevalent (date of data extraction 4 May 2022). Data cleaning, visualisation and analysis were conducted using Stata version 17 (StataCorp, TX, USA).

5.3 RESULTS

5.3.1 PHDC ascertained CKD

As of 31 December 2021, 89,542 healthcare clients known to the PHDC had evidence of CKD. This included those ascertained in 2010 up until the end of 2021. The number of new CKD episodes ascertained rose sharply after 2010 and dropped precipitously in 2020 and 2021 (Figure 5. 2). Figure 5. 3 shows that there were many individuals who disengaged with the healthcare system over time, especially after 2019. There was also substantially less eGFR testing in those with CKD during 2020 and 2021 compared to earlier years (Table 5. 1). Characteristics of those prevalent in 2019 by evidence of no prior death and evidence of recent engagement are summarised in Table 5. 1. Only the former derived group who had CKD ascertained up to and including 2019 who had no evidence of death is described here on.

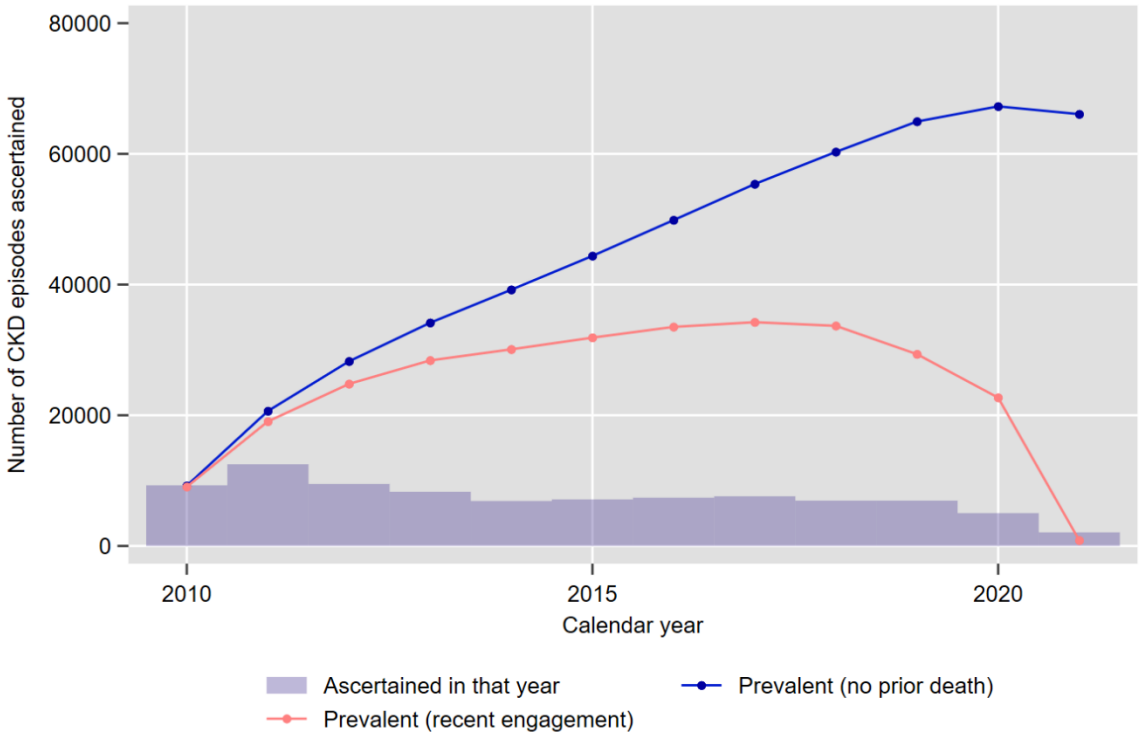


Figure 5. 2 CKD prevalent count trend over time using death and last healthcare engagement

The histogram shows newly ascertained CKD per calendar year and the line graphs show the cumulative prevalent CKD cases per year over time by two types of evidence: i) no prior death and so assumed alive and ii) engagement after the index year confirming survivorship. Abbreviations: **CKD**, chronic kidney disease.

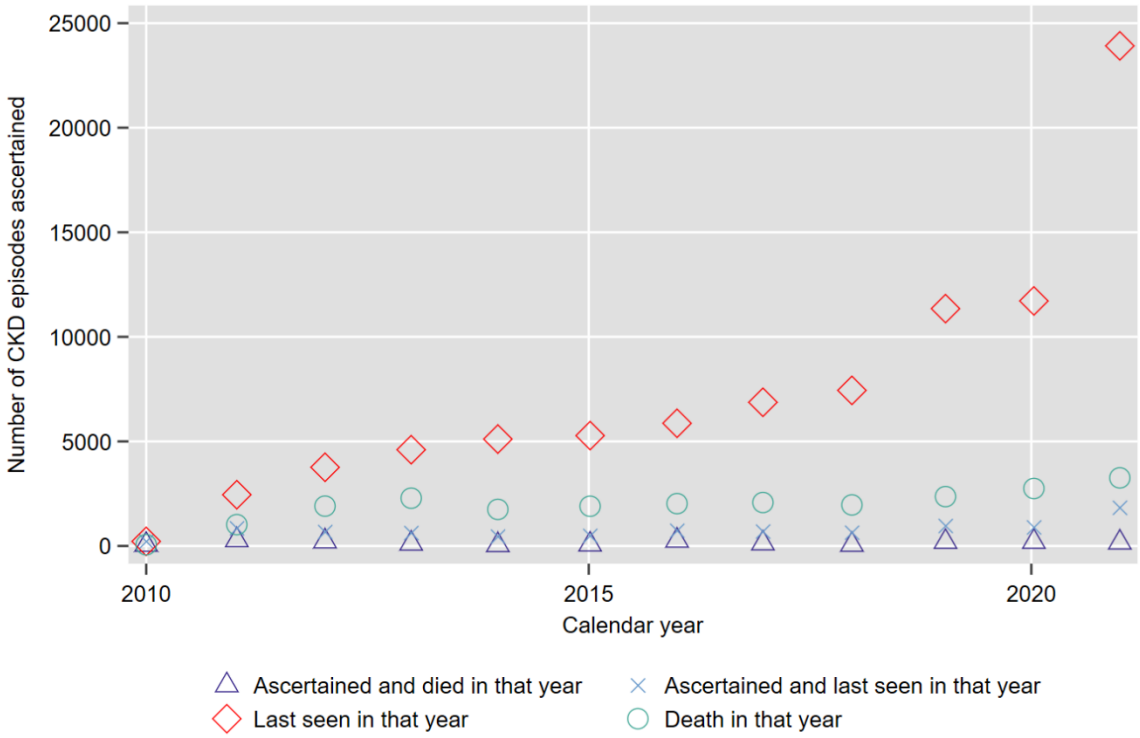


Figure 5. 3 Censoring reasons (death and healthcare disengagement) resulting in derived prevalent CKD counts

For each calendar year, the number of individuals with CKD ascertained with evidence of death or no evidence of future healthcare engagement is shown. In-facility death only explained a small proportion of why individuals were not seen again. Factors relating to the coronavirus-19 pandemic, out-of-facility deaths and inadequate follow-up time accrual before data extraction may be responsible for the pattern seen. Abbreviations: **CKD**, chronic kidney disease.

	Prevalent by non-death, N = 64,952	**Prevalent by evidence of recent engagement**, N = 29,351
Age category, %		
18 – 39	6.7%	7.2%
40 – 64	43.4%	46.1%
65 – 74	28.8%	29.7%
≥75	21.1%	17.0%
Mean age (SD)	63.5 (13.9)	63.9 (13.5)
Female sex, n (%)	63.9%	64.3%
Evidence for CKD, % yes (may have more than one)		
Two consecutive eGFR < 60 ml/min/1.73m^2	64,674 (99.6%)	29,311 (99.9%)
Dialysis ICD-9/10 code	514 (0.8%)	388 (1.3%)
Possible HD	1,354 (2.1%)	963 (3.3%)
Possible PD	1,097 (1.7%)	814 (2.8%)
Transplant procedure	416 (0.6%)	253 (0.9%)
Transplant medication	138 (0.2%)	111 (0.4%)
KDIGO eGFR CKD severity stage, n (%) *		
45 – 59.9	18,116 (68.7%)	11,179 (68.8%)
30 – 44.9	5,585 (21.2%)	3,545 (21.8%)
15.0 – 29.9	1,817 (6.9%)	1,092 (6.7%)
<15 ml/min/1.73m^2	757 (2.9%)	396 (2.4%)
eGFR testing patterns§		
N results per year		
2016	80,075	38,375
2017	83,324	42,805
2018	77,841	43,893
2019	80,961	51,441
2020	55,909	47,990
2021	53,113	46,134
N results per person		
1	3,504	554
2 – 5	23,880	9,941
6 – 10	17,052	10,490
> 10	11,628	8,006
Visit type at CKD ascertainment, %		
Primary Healthcare	53.4%	55.5%
Hospital outpatient	37.3%	35.6%
Admission	9.3%	8.9%

Facility type at CKD ascertainment, %	53.3%	55.7%
Primary healthcare	21.3%	20.4%
District Hospital	2.7%	2.4%
Regional Hospital	22.1%	21.1%
Central Hospital	0.6%	0.4%
Other^		
Comorbidities, yes %		
Hypertension	78.4%	83.0%
Diabetes	45.4%	48.4%
HIV	10.0%	11.3%

Table 5. 1 Characteristics of prevalent individuals with CKD in 2019 at ascertainment

* newly ascertained in 2019. ^ Other facility types include maternity obstetric units, oral health, psychiatric and tuberculosis hospitals, intermediate care, and satellite clinics. Abbreviations: **CKD**, chronic kidney disease; **eGFR**, estimated glomerular filtration rate; **KDIGO**, Kidney Disease: Improving Global Outcomes; **HIV**, human immunodeficiency virus.

Individuals' demographics and comorbidities are summarised in Table 5. 1. The mean age was 63.5 years old (SD 13.9) and most were female (63.9%). A substantial proportion had co-existing hypertension (78.4%) and diabetes (45.4%). Only 10.6% did not have hypertension, diabetes or HIV comorbidities, whereas 47.3% had ≥ two comorbidities.

Estimated GFR characteristics and testing: In the year preceding CKD ascertainment (newly ascertained CKD in 2019), the mean eGFR was 46.6 (SD 27.1) ml/min/1.73m^2. The average eGFR at CKD ascertainment was 47.2 (SD 12.0) ml/min/1.73m^2. After CKD ascertainment, the mean eGFR was 39.9 (SD 21.2) ml/min/1.73m^2 in the year that followed. eGFR values during episodes of acute-on-chronic kidney disease[15] were disregarded as these values do not reflect steady-state kidney function.

Healthcare encounters and eGFR testing: 66.8% of healthcare facility visits prior to CKD ascertainment were in primary care. Also, confirmatory eGFR tests that were defining of CKD were mostly undertaken at primary levels of care (53.4%), tertiary-level (22.1%) and district-level (21.3%) hospitals. In terms of visit types, 53.4% of CKD-defining eGFR tests were undertaken at the time of primary healthcare visits, 37.3% at hospital outpatient visits and 9.3% at hospital admission. Of the hospital admissions on the same day that CKD was ascertained, 29.7% were emergency medicine admissions and of hospital outpatient visits, 27.2% were ascertained at the emergency centre. There was a wide variation in the regularity

[15] Acute-on-chronic kidney disease is described in chapter 8.

of eGFR testing before ascertainment, median 8 days (IQR 2, 55) apart. After CKD ascertainment, testing frequency averaged every 56 days (IQR 4, 224).

5.3.2 Kidney replacement therapy

Of all prevalent healthcare clients with CKD in 2019, 1,755 had evidence of either procedure-coded dialysis, possible HD, possible PD, transplantation (procedure codes or immunosuppressive medication). The proportion of people with a dialysis ICD-9 procedure code who also had other evidence of KRT was 446/514 for possible HD, 429/514 for possible PD, 170/514 for transplantation procedure and 65/514 for transplantation medication.

The mean age of healthcare clients with any evidence of KRT was 43.0 years (SD 14.1) and about half (49.6%) were female. The average duration as of the date of last encounter was 6.3 years (SD 3.2)[16] from CKD ascertainment.

[16] The date of dialysis initiation was unknown and so it was not possible to calculate dialysis vintage.

5.4 DISCUSSION

Using an eGFR-based definition of CKD and inferred evidence of KRT recorded by the PHDC, the number and characteristics of people with CKD accessing healthcare in the City of Cape Town were described here. Individuals were most frequently young, had stage IIIa CKD and comorbidities were very common.

5.4.1 Chronic kidney disease burden and features

After a period of stabilisation, CKD ascertainment dropped significantly in 2020 and 2021. This coincides with the COVID-19 pandemic so may be due to decreased eGFR testing in general. Limited testing may also have led to a lack of confirmatory eGFR below 60ml/min/1.73m^2 by the end of the data extraction period. More seriously, the substantial reduction in patient engagement in 2020 and 2021 is worrisome as a harbinger of lost opportunity to modify risk factors and monitor and treat complications of CKD including progression.

Although only counts of incident and prevalent CKD were possible in this analysis, two population based cohort studies in the City of Cape Town, mostly of mixed ancestry individuals, would suggest a stage 3 – 5 CKD age-adjusted prevalence of between 6.4% (95% CI 3.2, 9.7%) and 8.7% (95% CI 5.0, 8.5).(203,206) Peer et al highlighted that kidney dysfunction (single eGFR < 60ml/min/1.73m^2) in black township residents in the City of Cape Town differed by eGFR equation used: between 2.0% for Cockcroft-Gault to 5.9% by CKD-EPI$_{cr-cys}$.(207) By MDRD equation without the race coefficient, as received by the PHDC, calculated frequency was 5.1% in that study. A meta-analysis which found that using the CKD-EPI$_{cr}$ 2009 compared to the MDRD equation resulted in fewer CKD cases, but classification of those who

experienced kidney failure (KF) and mortality was more accurate.(208) Nevertheless, eGFR equations do not perform the same in Africa and even the refit CKD-EPI 2021 equation is not recommended for use in Africa.(86)

The demographics of individuals with evidence of CKD highlights a young patient population. Despite this, the cardiovascular risk factors, diabetes and hypertension, were common in those with CKD. The South African Renal Registry 2020 report indicated that the cause for KF in patients receiving long-term KRT was hypertensive kidney disease in 36.6% and diabetic kidney disease in 14.3%.(195) Although the PHDC does not collate information on the specific cause of kidney disease, hypertension and diabetes are likely to be leading causes and consequences of those with CKD in the City and tighter control of these risk factors are necessary to prevent the development and progression of CKD.

Concurrent HIV infection in the analysed healthcare clients with CKD was slightly higher than in the general population (of the City of Cape Town).(191) The proportion of healthcare clients who were HIV positive was lower than expected given the fact that the PHDC is enriched with persons accessing HIV care.(192) Nevertheless, a PHDC analysis of HIV-negative versus -positive healthcare clients in the Khayelitsha sub-district of Cape Town, found that the age at CKD ascertainment to be on average 16 years younger for PLHIV suggesting either HIV is a risk factor for earlier CKD development or demonstrates the improved surveillance of CKD in PLHIV as a result of regular kidney function testing in this population.(174)

Although kidney function testing leading to the ascertainment of CKD often coincided with primary healthcare visits, it is unknown how well CKD is recognised by treating clinicians. A web-based health dashboard ('Single Patient Viewer') still in pilot phase in the province, will make important CKD metrics readily available to clinicians which should serve as a prompt to recognition and allow convenient monitoring of CKD severity and progression.(80)

5.4.2 Kidney replacement therapy

Dialysis units voluntarily send information on new and existing patients receiving KRT care to the South African Renal Registry.(195) According to the 2020 report, there were 1,877 patients on KRT in the Western Cape, 843 of which were being treated at public healthcare facilities, with a KRT prevalence of 150 per million population. Two adult public healthcare sector dialysis centres offer KRT services in the City: Tygerberg Hospital (244 patients) and Groote Schuur Hospital – 484 patients (personal communication with Prof M Razeen Davids of the South African Renal Registry). The PHDC alone enumerated many more with coded dialysis, possible HD and possible PD in the City indicating an overestimation of KRT by the PHDC (compare 244 + 484 to the 1,755 enumerated by the PHDC). Although submission to the registry is voluntary, the dialysis centres included in this analysis are the province's two academic institutions, one of which where the sitting registry's chief is employed. There may be clinical or coder misuse of diagnostic and procedure codes, or a low specificity of evidence used by the PHDC to identify KRT. The PHDC may also capture patients undergoing temporary dialysis pending assessment for the long-term KRT program, some of whom will be declined and so not be reported to the Registry. In addition, there may also be under-ascertainment of kidney transplantation: according to organ donor data: there were 2,346 transplants between 1991 to 2015 in the Western Cape public sector, although some recipients may have since died

or were not domiciliary in the Western Cape.(209) Further validation work is underway to improve the specificity and sensitivity of KRT evidence used by the inference engine of the PHDC.

5.4.3 Strengths and limitations

The data the PHDC receives and who accesses public healthcare: The systematic capture of CKD using routinely collected data deserves mention. The PHDC collects and warehouses data from public healthcare facilities in the Western Cape, and primary, secondary and tertiary levels of care data are integrated with laboratory, medication, encounter and other clinical information in near real-time using a universal patient identifier and updated longitudinally.(79)

Drawbacks are that the PHDC is unable to track people accessing private and out-of-province healthcare providers. This is pertinent to CKD given that KRT access is limited in South Africa and people not able to access KRT in the public healthcare system may have no other choice than to migrate back to rural family homesteads, usually outside of the province, because they cannot afford private care or may pay out-of-pocket in the private sector.(210) Although only a small proportion of people had evidence of KF (eGFR < 15 ml/min/1.73m^2) at ascertainment, the results confirm the substantial need for KRT in the public sector. Also, it is worth noting that Groote Schuur Hospital is one of a few national transplant centres for the public healthcare sector. Transplantation numbers may therefore have been over-enumerated by transplant recipients from outside of the province.

Consider also that the City of Cape Town is an urban area, interspersed with informal housing. It is the residents of these informal areas that often access public healthcare facilities from which the PHDC receives data. The demographics and severity of CKD described here are therefore not generalisable to the whole City or Province. Because specialist nephrology services are concentrated in Cape Town, CKD ascertained outside of the City may be included in the described results if referred, particularly for assessment for KRT. The fact that testing leading to CKD ascertainment appears to be de-centralised to primary healthcare providers would suggest that it is only advanced CKD that may be over-represented in this analysis. Furthermore, CKD ascertained through testing done at central hospital level (tertiary/ quaternary level of care) should be interpreted in light of the fact that these hospitals occasionally function as secondary-level referral centres in sub-districts where there are no district-level hospitals. For example, Groote Schuur Hospital accepts primary healthcare and self-referrals from its neighbouring suburbs. Hospitals also receive referrals from private primary healthcare general practitioners and so CKD may erroneously appear to be ascertained at hospital-level because of how people access a complex healthcare system in South Africa.

Challenges in ascertainment: Inaccuracies in ascertainment may also have been introduced due to duplication of unique patient identifiers. Information on the same person is possibly received by the PHDC under multiple distinct patient identifiers. Encounter, SCr and other data used to define CKD and KRT may thus fail to be consolidated across duplicated records. The PHDC estimates duplication to affect 9.6 – 16.8% of the patient master index and an extensive amount of work is ongoing to improve record linkage between possible related

duplicates.(176) Further bias would be introduced by defining the start of CKD as the first eGFR < 60ml/min/1.73m^2 because the individual would need to survive, or reside in the City, long enough to undergo confirmatory testing (lead time bias). Death, or disengagement with the healthcare system, may occur as a direct consequence of CKD (competing risk) and CKD may erroneously be confirmed by an AKI episode (misclassification bias). These issues are well known when describing kidney disease epidemiology with routinely collected data.(211,212)

The addition of albuminuria to the definition and staging of PHDC ascertained CKD would be advantageous in the future to not only fully align with KDIGO criteria but also improve the accuracy of detection and staging given the limitations of SCr based eGFR equations.(26) Also to be considered, is the addition of the measurement of Cystatin-C which is less affected by non-GFR factors and mirrors mGFR closer than both the 4- variable MDRD and CKD-EPI$_{cr}$ in Africans.(86)

Finally, reduced eGFR testing and patient engagement during 2020 – 2021 due to the restrictions imposed as a result of the COVID-19 pandemic may have impacted the number of CKD ascertained detected during that time.

5.5 CONCLUSION

Routinely collected data were used to describe the clinical epidemiology of CKD in the City of Cape Town. A description of the clinical characteristics and burden of kidney disease was novel and important given the prevalence of these diseases. While the utilisation of the health information exchange that aims to centralise and consolidate administrative and clinical information is promising and unique to the province, this analysis highlighted important challenges in the use of rule-based algorithms, that infer conditions using multiple types of evidence, to detect kidney diseases in the context of the complex South African healthcare system and imperfect kidney function measures. The following chapter explores the incidence and characteristics of person with acute-on-chronic kidney disease in the City of Cape Town.

6 INCIDENCE AND CHARACTERISTICS OF HEALTHCARE CLIENTS WITH ACUTE-ON-CHRONIC KIDNEY DISEASE IN THE CITY OF CAPE TOWN, SOUTH AFRICA

6.1 INTRODUCTION

AKI in people with CKD is common and may result in accelerated kidney disease progression.(43,213) Acute-on-chronic kidney disease (A-on-CKD) has not frequently been studied in detail, despite knowledge of how acute and chronic kidney diseases intersect and the fact that both are separately highly prevalent in South Africa.(42) The PHDC assigns kidney function as one of four possible states: normal, single abnormal SCr, AKI or CKD (section 1.4.11). Given that the PHDC assigns a status of AKI or CKD as mutually exclusive disease episodes, an algorithm adapted from the NHSE detection algorithm, was used to ascertain AKI in people with established CKD (*PHDC-ascertained CKD*).(63) In this chapter, the CKD population in which A-on-CKD was detected is described as well as the amendments to the NHSE algorithm made. The results of the analysis and a discussion follow later in the chapter.

6.2 METHODS

6.2.1 Study design

This analysis used prospective longitudinal SCr data collected from a cohort of healthcare clients with PHDC-ascertained CKD to describe A-on-CKD.

6.2.2 Study population

A-on-CKD was detected in adult (≥ 18 years old) healthcare clients with PHDC-ascertained CKD. Data on SCr were extracted for such individuals. Only SCr data for 2016 – 2021 were available for this analysis. Thus, the algorithm was deployed during 2017 – 2021 with 2016 reserved for looking back in the preceding year to establish the baseline SCr. Individuals with any evidence of KRT were excluded based on possible HD, possible PD, dialysis procedure codes, transplantation medication and transplantation procedure codes. Because the date of KRT initiation is unknown, such persons were excluded entirely for the purpose of this analysis as changes in SCr before/ after KRT may be misclassified as AKI.

6.2.3 Acute-on-chronic kidney disease detection algorithm

Several modifications to the NHSE algorithm were enforced because the time of specimen collection was not captured, and theoretical deficiencies of the algorithm were identified in other parts of this thesis (section 3.2.5). These modifications and their justifications are detailed below. The Stata code is made publicly available at https://github.com/RyAylwd/AKIalgorithm.

Firstly, the 48-hour rule was replaced with a 2-day rule by date in the case of KDIGO stage 1 criteria as only the date, and not the time, of specimen collection was available.(11) Secondly, the NHSE algorithm updates the baseline SCr used to trigger future AKI alerts, with values from during the current AKI episode potentially inflating the baseline during an AKI episode, making it more difficult to trigger subsequent AKI alerts. As such, in this analysis, the baseline SCr was anchored to the baseline value that triggered the AKI in the first instance for the duration of the current episode. Thirdly, the SCr result immediately following one AKI episode

was reset as the new baseline to which subsequent results were compared. In other words, all previous SCr results were disregarded. This acknowledges the fact that the SCr following an insult may not recover to the previous baseline but, on the other hand, might be falsely low due to non-GFR factors that affect SCr levels. The assessment of the baseline look-back period is summarised in Figure 6. 1.

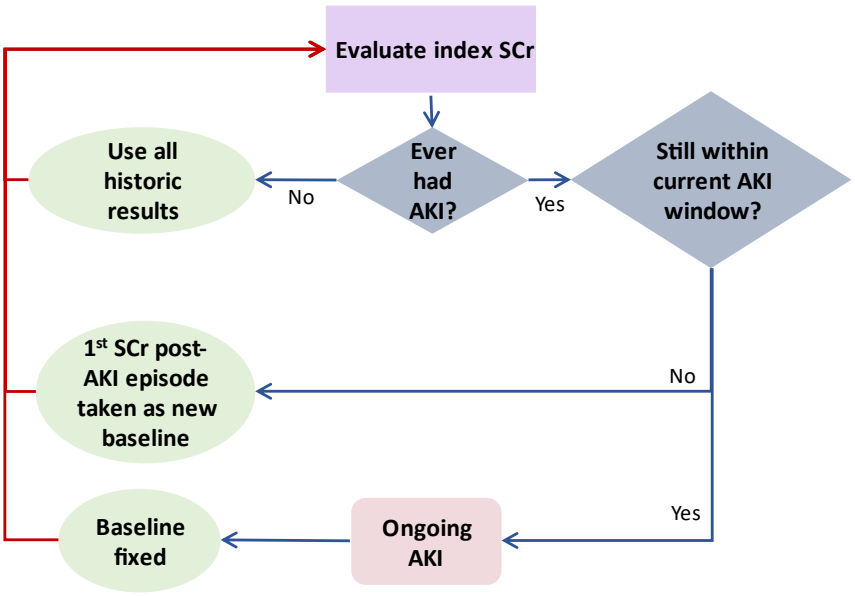


Figure 6. 1 Selection of the baseline assessment look-back period

Each individual SCr result was evaluated in turn. Depending on whether or not AKI had previously been experienced or whether the AKI was still currently active, the baseline SCr was either fixed or reset if an AKI episode was still active or previously closed, respectively. Abbreviations: **AKI**, acute kidney injury; **SCr**, serum creatinine.

Fourthly, in keeping with the KDIGO definition of AKI as injury lasting ≤ 7 days, the AKI episode was automatically closed after 7 days.(10,35) Fifthly, an assessment of recovery was included, which was designated 'complete', 'partial', 'non-recovery' or 'uncaptured' using the rules shown in Table 6. 1.(214) The most recent SCr value on or before the end of the 7-day AKI episode was compared to the baseline SCr value to determine recovery. KDIGO criteria of

recovery of an eGFR > 60ml/min/1.73m^2 was not used as patients with CKD would never meet such criteria as they had a starting eGFR < 60ml/min/1.73m^2.(11) The possible outcomes of AKI are illustrated conceptually in Figure 6. 2. Figure 6. 3 summarises how the algorithm was deployed in the longitudinal SCr data and Figure 6. 4 summarises the relationship between the first and subsequent AKI episodes and how their baseline SCr and recovery were assessed.

Recovery status	**Criteria**
Complete	≤ 20%
Partial	> 20% & < 50%
Non-recovery	≥ 50%
Uncaptured	No SCr result available after AKI start to assess future recovery

Table 6. 1 Recovery assessment used in the modified NHSE algorithm

Complete, partial and non-recovery was declared if the most recent SCr within the 7 days from AKI start was ≤ 20%, > 20% but < 50% or ≥50% of the baseline SCr value. Abbreviations: **AKI**, acute kidney injury; **SCr**, serum creatinine.

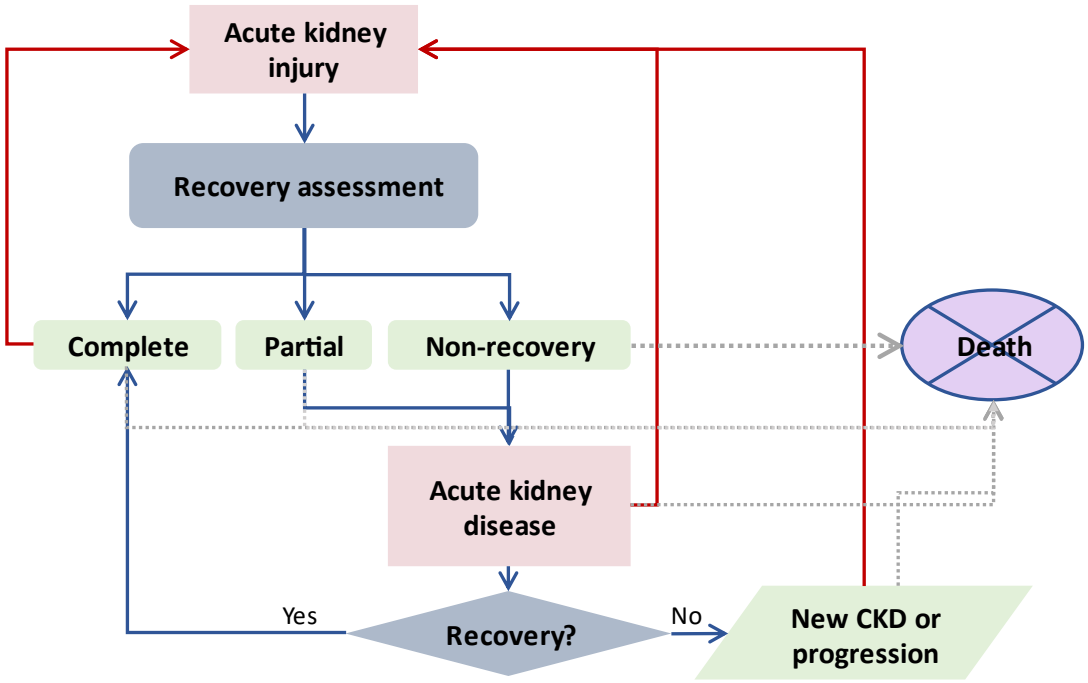


Figure 6. 2 Potential immediate or long-term outcomes following acute kidney injury

Following an AKI episode, kidney function recovery may be complete, partial or there may be non-recovery. In the continuum, partial or non-recovered AKI would be called AKD, and non-recovery of AKD would be called CKD. Death may complicate all forms of kidney disease independent of kidney function recovery. Abbreviations: **AKI**, acute kidney injury; **AKD**, acute kidney disease; **CKD**, chronic kidney disease.

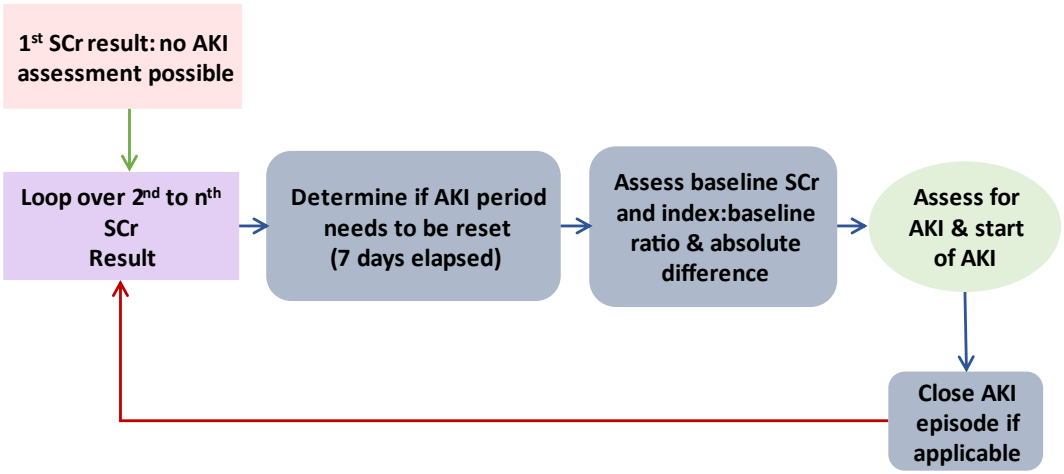


Figure 6. 3 Algorithm workflow used to detect AKI

This flowchart shows the procedure for detecting AKI by executing the detection code iteratively over longitudinal SCr measurements. AKI could not be detected using the individual's first ever SCr result as no baseline SCr was recorded before this time. Subsequent SCr results were evaluated sequentially in a loop that assesses firstly whether there was an open AKI episode, which would then inform the decision to reset the baseline SCr or continue to fix it for the duration of the still active AKI. Next, the AKI stage (0, 1, 2 or 3) was determined using the index SCr value compared to the baseline SCr (based on KDIGO criteria). Lastly, if 7 days had elapsed since the start of the AKI, then the episode was closed automatically. Once completed, the steps looped over the next future SCr result. Abbreviations: **AKI**, acute kidney injury; **SCr**, serum creatinine; **KDIGO**: Kidney Diseases: Improving Global Outcomes.

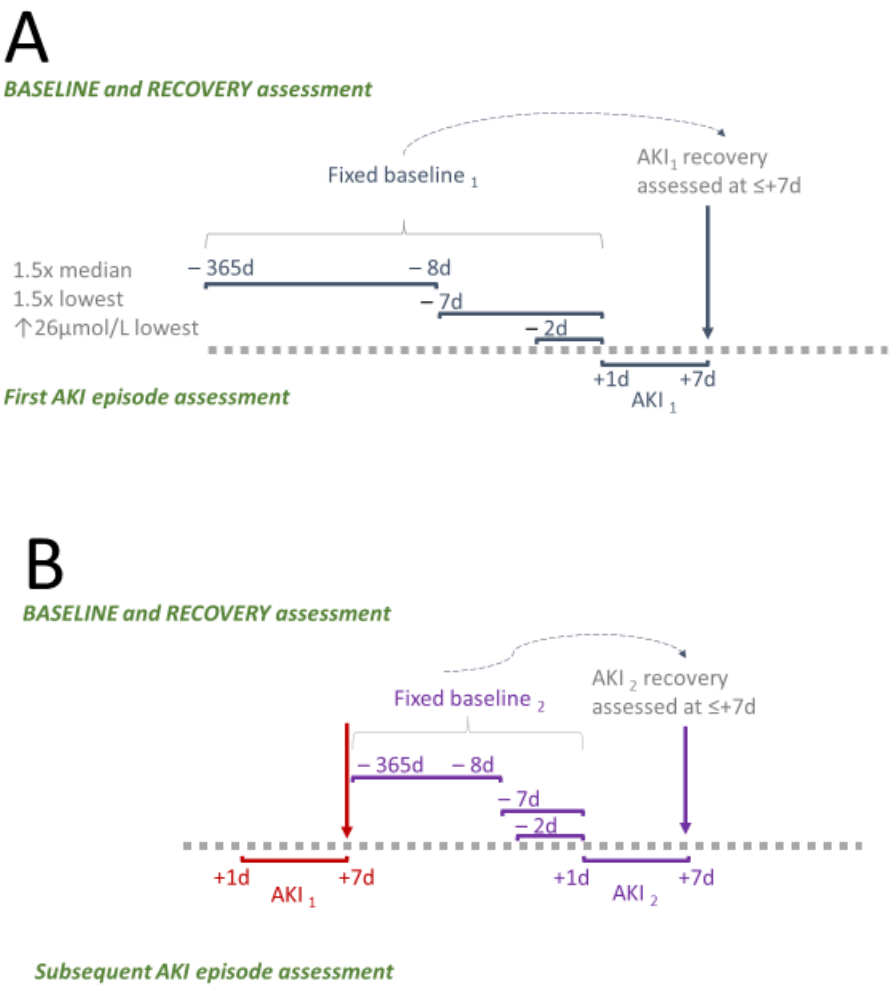


Figure 6. 4 Assessment of the AKI episode, baseline and recovery

These diagrams summarise the different assessment procedure for the first AKI episode (A) and subsequent AKI episodes (B). The dotted line represents SCr values. Shown above the dotted line, available SCr results are assessed in the preceding 2 days, 7 days and 365 days to inform the most appropriate baseline. The lowest value from the preceding 2 and 7 days is compared to the median value of all SCr results in the preceding 8 – 365 days. The higher of the baseline SCr: index SCr ratios triggers AKI, if appropriate. The baseline SCr is anchored for the entire AKI episode, which is automatically closed after 7 days of the start of the episode. The same baseline SCr is used to assess kidney function recovery status. For subsequent AKI episodes (B), a new baseline period is established before AKI can be triggered again, but values are only regarded after the previous AKI episode and up until 365 days before the subsequent AKI episode. Abbreviations: **AKI**, acute kidney injury; **AKD**, acute kidney disease; **d**, day.

6.2.4 Analysis

Appropriate summary statistics (counts, proportions, mean [SD], median [IQR]) were used to compare those who did and did not experience A-on-CKD.

The detection of A-on-CKD was definitive: an individual with PHDC-ascertained CKD had SCr results available, and the executed algorithm detected changes in SCr consistent with AKI. What was uncertain was who did not experience AKI. To potentially detect A-on-CKD, persons with ascertained CKD would need to be alive. Only in-facility deaths are reliably recorded by the PHDC so out-of-facility deaths would not be known. As such, survivorship was assumed if an individual had no previous recorded death before 1 January 2017 and had evidence of a healthcare facility encounter in or after 2017.

The A-on-CKD incidence rate and cumulative incidence were described. The incidence rate is the number of new A-on-CKD cases (numerator) divided by the total person-time of all persons who were at risk of developing A-on-CKD (denominator). In contrast, the cumulative incidence (risk) was calculated as the number of persons who experienced A-on-CKD divided by the number of persons who were at risk of developing A-on-CKD during the study period. Although recurrent AKI episodes were possible, only the first episode was regarded. The incidence should thus be interpreted as the rate or risk of *at least one* A-on-CKD episode. Person-time was calculated using the time from when the person became at-risk until the person was no longer at-risk of developing the first episode of algorithm-detected A-on-CKD. Time at-risk for A-on-CKD began on the date of CKD ascertainment if CKD was ascertained 2017 – 2021 or on 1 January 2017, when the AKI algorithm was activated, for CKD ascertained before 2017. Time at-risk ended on the date of in-facility death, if recorded, the date of last

facility engagement or the date of the first AKI episode start, whichever came first. Due to the unavailability of SCr results prior to 2017, the detection of A-on-CKD in individuals with CKD ascertained before 2017 may not have been the person's first ever episode. As such, incidence was described for those with incident CKD ascertained on or after 1 January 2017 and separately for those with ascertained CKD before 1 January 2017. In addition, the incidence rate and cumulative incidence were stratified by sex. This was pertinent given the fact that the PHDC data are enriched with women accessing routine contraceptive and maternal care and so the incidence of AKI may be different for non-biological reasons.

Furthermore, the characteristics of persons with community- (CA-AKI) and hospital-acquired AKI (HA-AKI), including severity by KDIGO staging and recovery, were described. HA-AKI was defined as occurring on day three or later of a facility admission and hospitalised CA-AKI was defined as occurring when an AKI episode occurred within 7 days inclusive before the start of an admission and 2 days into an admission. Only the first episode was described in detail for ease of interpretation. Patient encounters with the healthcare system 30 days before and at the time of the AKI episode were illustrated using Venn diagrams.

In-facility mortality, within the index admission, following hospitalised CA-AKI and HA-AKI was visualised using bar charts stratified by age category. Also, since AKI recovery statuses 'non-recovery' and 'uncaptured' were possibly dependent on whether the individual died and an individual with 'complete' or 'partial' recovery status may still have subsequently died, death was stratified by recovery status. In-hospital death was assumed if the death was in the same month as the recorded A-on-CKD episode. Mortality was not described for persons without A-on-CKD as only in-facility deaths were recorded and the focus of this analysis was outcomes

following AKI events and not the sequelae of all hospital admissions experienced by persons with CKD.

Data cleaning, visualisation and analysis were conducted using Stata version 17 (StataCorp, TX). Incidence rates of A-on-CKD were calculated using the `stptime` command.

6.3 RESULTS

A flow diagram of inclusion/exclusion criteria for the analysis of A-on-CKD is provided in Figure 6. 5. Out of the remaining healthcare clients with non-KRT CKD with evidence of being alive after 2017, there were 19,701 A-on-CKD episodes affecting 11,798 individuals (mean age 66.6 years [SD 14.0], 61.8% female). Most experienced a single AKI episode (1 episode: 70.3%, 2 episodes: 21.0%, ≥3 episodes: 8.7%). The median time until a subsequent AKI episode was 195 days (IQR 55, 437). There was a median of 8 (IQR 5, 14) SCr tests available for each individual to permit AKI detection.

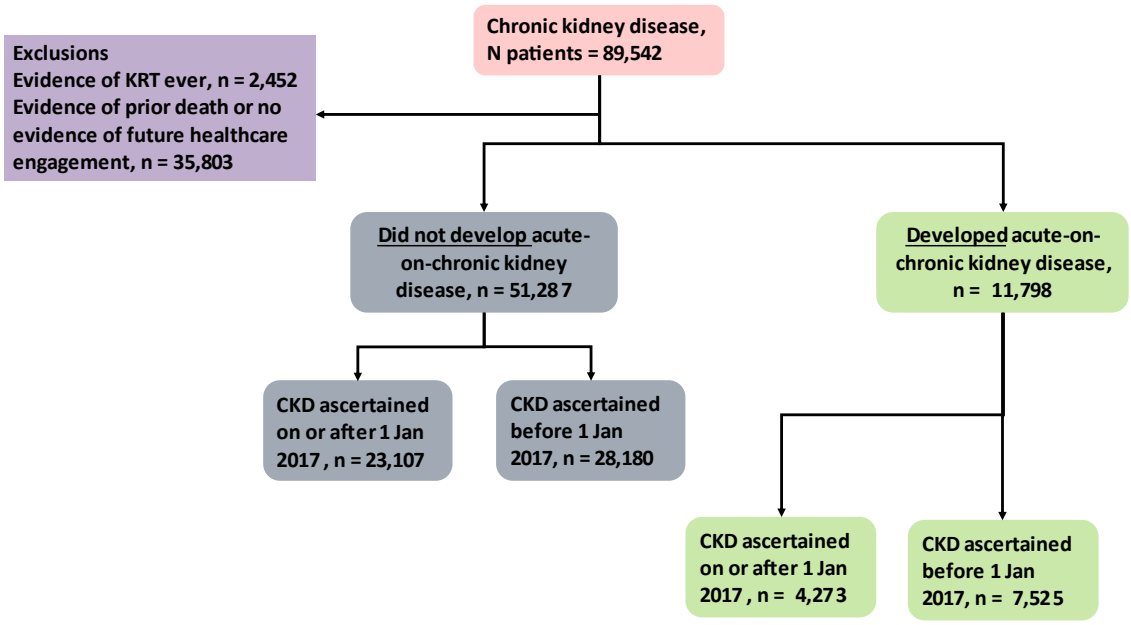


Figure 6. 5 Inclusion of health-seekers with and without acute-on-chronic kidney disease 2017 – 2021

The number of those with PHDC-ascertained CKD and at-risk of acute-on-chronic kidney disease are shown. Those who did and did not develop AKI are differentiated by whether or not they had incident or prevalent CKD at the time they were at-risk of an episode. Abbreviations: **AKI**, acute kidney injury; **KRT**, kidney replacement therapy.

6.3.1 Comparison of those with and without acute-on-chronic kidney disease

See Table 6. 2 for a comparison of demographics and comorbidities between those who did and did not experience AKI.

	Did not develop AKI		***Developed AKI***	
	Prevalent on 1 January 2017, n = 28,180	**Incident on or after 1 January 2017 ,** n = 23,107	**Prevalent on 1 January 2017,** n = 7,525	**Incident on or after 1 January 2017 ,** n = 4,273
Age category, %				
18 – 39	4.7%	7.4%	6.1%	8.1%
40 – 64	42.6%	48.0%	45.4%	46.8%
65 – 74	31.3%	26.8%	30.0%	27.4%
≥75	21.5%	17.9%	18.6%	17.7%
Age in years, mean (SD)	64.6 (12.8)	62.0 (13.8)	63.0 (13.2)	61.8 (14.1)
Sex				
Female, n (%)#	66.3%	62.7%	63.1%	59.3%
KDIGO eGFR CKD severity stage category*, %				
≤15 ml/min/1.73m^2		2.2%		4.9%
15.1 – 29.9		6.8%		9.4%
30 – 44.9		21.6%		22.3%
45 – 59.9		69.1%		63.2%
Pre-existing comorbidities, yes %				
Diabetes	43.7%	42.1%	59.6%	56.1%
Hypertension	81.6%	75.2%	85.1%	78.8%
HIV	7.8%	18.9%	10.5%	15.6%

Table 6. 2 Health-seeker attributes of those who did and did not develop acute-on-chronic kidney disease 2017 – 2021

Sex was missing for <1% of individuals. *staging at CKD ascertainment if ascertainment was in 2017 or later given the limited availability of eGFR results. Abbreviations: **AKI**, acute kidney injury; **IQR**, interquartile range; **KDIGO**, Kidney Diseases: Improving Global Outcome; **eGFR**, estimated glomerular filtration rate; **CKD**, chronic kidney disease; **HIV**, human immunodeficiency virus.

Table 6. 3 shows the incidence rate and person-time stratified by sex and incident and prevalent CKD populations. The rate of A-on-CKD was consistently higher for males compared to females despite a female predominance. The rate of A-on-CKD was higher for healthcare clients with CKD ascertained in or after 2017 compared to those with CKD ascertained prior to 2017. The cumulative incidence of A-on-CKD during the period 2017 – 2021 was 4,662/27,769 (16.8%) for CKD ascertained after 2017 and 8,810/36,990 (23.8%) for CKD ascertained before 2017.

	Overall	**Male sex**	**Female sex**
Overall	*150,098 pyar*	*52,170 pyar*	*97,844 pyar*
	78.6 per 1000-pyar (95% CI 77.2, 80.0)	86.3 per 1000-pyar (95% CI 83.8, 88.8)	74.5 per 1000-pyar (95% CI 72.8, 76.2)
Incident on or after 1 January 2017	*49,538 pyar*	*18,301 pyar*	*31,206 pyar*
	86.3 per 1000-pyar (95% CI 83.7, 88.9)	95.0 per 1000-pyar (95% CI 90.6,99.5)	81.1 per 1000-pyar (95% CI 78.0, 84.4)
Prevalent on 1 January 2021	*100,559 pyar*	*33,869 pyar*	*66,638 pyar*
	74.8 per 1000-pyar (95% CI 73.1, 76.5)	81.5 per 1000-pyar (95% CI 78.6, 84.6)	71.4 per 1000-pyar (95% CI 69.4, 73.4)

Table 6. 3 Person-time at-risk and incidence rates of acute-on-chronic kidney disease stratified by sex and incident/ prevalent CKD population

Person-time and incidence rates of healthcare clients with CKD ascertained prior to 2017 with no evidence of death prior to 2017 and affirmative evidence of healthcare engagement after 2017 were compared to healthcare clients with CKD ascertained in or after 2017, stratified by sex. All rates are reported per 1000 person-years at-risk (**pyar**).

Baseline creatinine: The median baseline SCr, 2 days, 7 days and 365 days pre-AKI was 243 (IQR 175, 395), 227 (IQR 157, 370) and 127 (IQR 101, 180) µmol/L, respectively. The median baseline SCr that triggered the AKI was 126 µmol/L (IQR 99, 179).

Severity: during the AKI episode, the median SCr was 255 (IQR 186, 407) µmol/L, the peak was 273 (IQR 193, 453) µmol/L, and the nadir was 208 (IQR 144, 340) µmol/L. The KDIGO stage at presentation was KDIGO stage 1 for 61.4% of AKI, stage 2: 21.1% and stage 3: 17.6%. The highest KDIGO stage reached during the episode was stage 1 AKI: 54.5%, followed by stage 2: 22.7% and stage 3: 22.8%. Table 6. 4 shows the SCr value characteristics during AKI by peak stage of 1, 2 or 3.

Creatinine, median (IQR) µmol/L	**Peak stage**		
	1	**2**	**3**
Index	209 (165, 291)	264 (212, 350)	557 (341, 939)
Nadir	185 (134, 261)	205 (141, 291)	432 (216, 828)
Peak	211 (167, 296)	281 (227, 372)	661 (438, 1037)
Final	187 (136, 266)	217 (145, 307)	532 (264, 919)

Table 6. 4 Creatinine characteristics by peak KDIGO stage for individuals with acute-on-chronic kidney disease 2017 – 2021

Abbreviations: **KDIGO**, Kidney Diseases: Improving Global Outcomes; **IQR**, interquartile range; **AKI**, acute kidney injury.

6.3.2 Community- versus hospital-acquired acute-on-chronic kidney disease

Shown in Table 6. 5, the median SCr pre-AKI as well as during AKI, was higher for CA-AKI compared to HA-AKI potentially highlighting delayed presentation. Despite this, the proportion of AKI that was staged as 1, was similar. There was slightly more HA-AKI stage 2 than CA-AKI stage 2, but only slightly more frequent CA-AKI stage 3 than HA-AKI stage 3. Post-AKI, in the year following the end of AKI, the median SCr was higher for CA-AKI (224; IQR 137, 449 µmol/L) compared to HA-AKI (186; IQR 125, 315 µmol/L). Most individuals with CA-AKI was hospitalised within 7 days (65.6%).

	All AKI episodes, n episodes = 11,798	**Community acquired**, n episodes = 9,307	**Hospital acquired**, n episodes = 2,491
Age category, %			
18 – 39 years	4.8%	5.5%	2.9%
40 – 64	36.4%	37.9%	29.5%
65 – 74	29.0%	27.8%	33.8%
≥75	29.8%	28.8%	33.8%
Female sex, %	61.8%	61.7%	62.2%
Number of episodes, N (%)			
Per year			
2017	2,687 (21.6%)	2,218 (22.5%)	469 (18.6%)
2018	2,380 (20.2%)	1,923 (20.8%)	457 (18.4%)
2019	2,296 (20.0%)	1,804 (19.9%)	492 (20.1%)
2020	2,162 (18.7%)	1,644 (18.1%)	518 (20.9%)
2021	2,273 (19.5%)	1,718 (18.7%)	555 (22.0%)
Creatinine, median (IQR) µmol/L*			
Pre-AKI			
2 days	243 (175, 395)	255 (184, 417)	204 (147, 323)
7 days	227 (157, 370)	250 (179, 411)	150 (110, 230)
365 days	127 (101, 180)	127 (100, 184)	126 (102, 168)
At ascertainment	255 (186, 407)	261 (188, 422)	238 (181, 355)
Peak during episode	273 (193, 453)	276 (194, 463)	263 (191, 428)
Nadir during episode	208 (144, 340)	217 (149, 360)	187 (129, 287)
Final during episode	217 (147, 372)	223 (152, 386)	199 (133, 331)
AKI KDIGO peak severity stage, %			
1	54.5%	54.5%	54.5%
2	22.7%	22.1%	24.7%
3	22.8%	23.4%	20.9%
Comorbidities*#, yes %			
Hypertension	81.5%	82.1%	80.4%
HIV	12.0%	12.9%	7.6%
Diabetes	56.5%	56.5%	58.0%

Table 6. 5 Characteristics of the first episode of acute-on-chronic kidney disease 2017 – 2021

Abbreviations: **N**, number; **IQR**, interquartile range; **AKI**, acute kidney injury; **KDIGO**, Kidney Diseases: Improving Global Outcomes; **HIV**, Human Immunodeficiency Virus.

Healthcare facility encounters in relation to AKI episodes: At the time of acute-on-chronic disease, 28.6% of hospital admissions were admitted via emergency medicine[17], and 44.6% of hospital outpatient visits were emergency centre visits. Individuals with hospitalised CA-AKI were most frequently treated in district hospitals (56.3%) and central hospitals (33.7%); the median length of hospital stay was 4 days (IQR 2, 8). For HA-AKI, admissions were most frequently to a central hospital (48.9%) and district hospital (42.8%) and the median length of hospital stay was 11 days (IQR 8, 18). Facility encounters in the 30 days before episodes of A-on-CKD are presented in Figure 6. 6 (CA-AKI) and Figure 6. 7 (HA-AKI).

[17] Patients may have been referred in directly via a medical speciality thus bypassing an emergency medicine specialty or emergency centre visit record. Also, some disciplines (obstetrics and gynaecology) are separate from the emergency centre and were not included as emergencies. Hence, emergency admissions here may not include all emergencies depending on how the individual was processed by the facility.

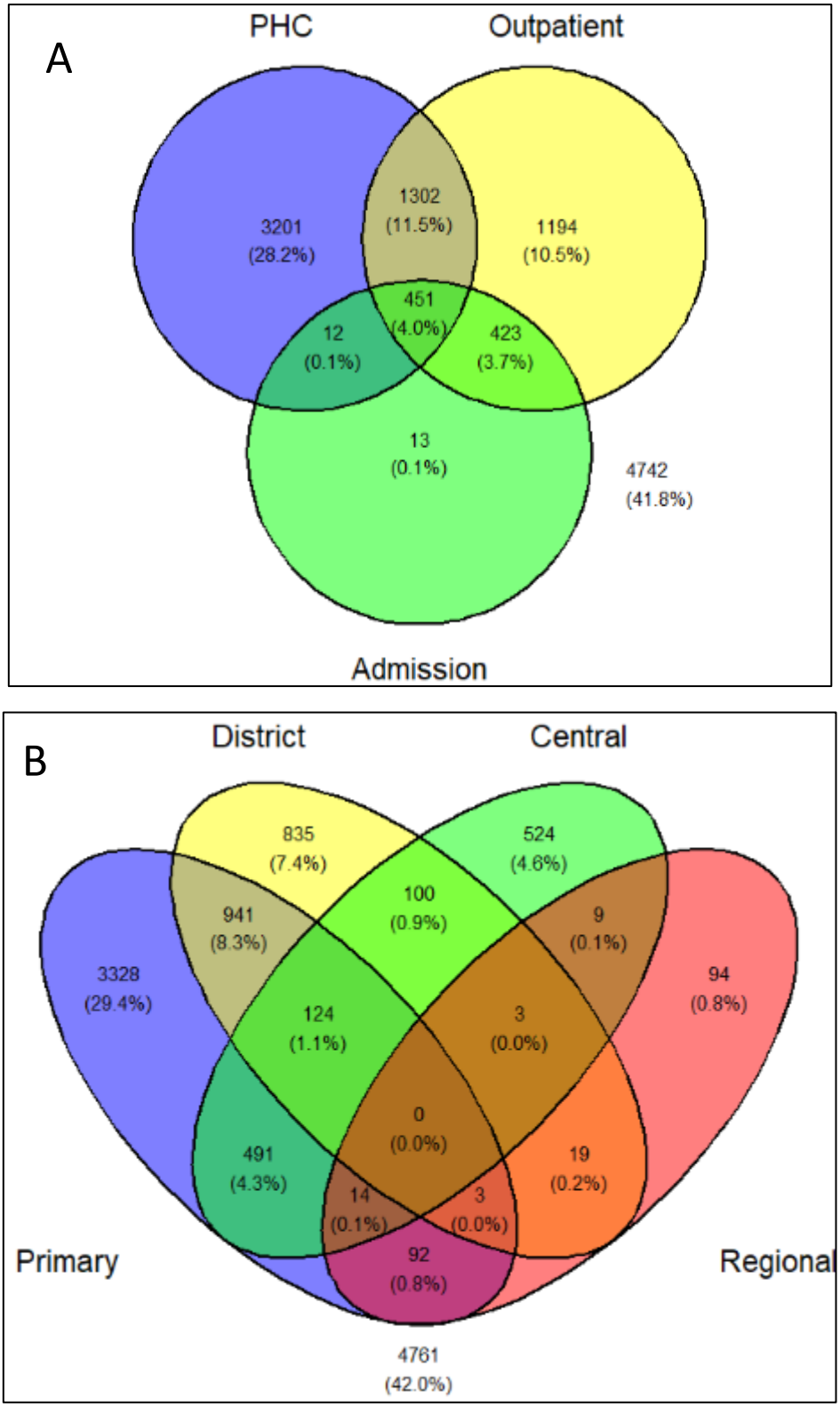


Figure 6. 6 Healthcare visit and facility types in the 30 days prior to community-acquired acute-on-chronic kidney disease

Intersecting circles are shown for (A) the number of primary healthcare (PHC), hospital outpatient and hospital admission visits and (B) primary healthcare, district hospital, regional hospital and central hospital facilities healthcare clients engaged with in the 30 days prior to their index community-acquired acute-on-chronic kidney disease episode. The number of the remaining individuals with no evidence of any healthcare engagement prior to their index CA-AKI episode are shown outside of the diagrams. Drawn using the R ggvenn package. Abbreviations: **CA-AKI**, community-acquired acute kidney injury; **PHC**, primary healthcare.

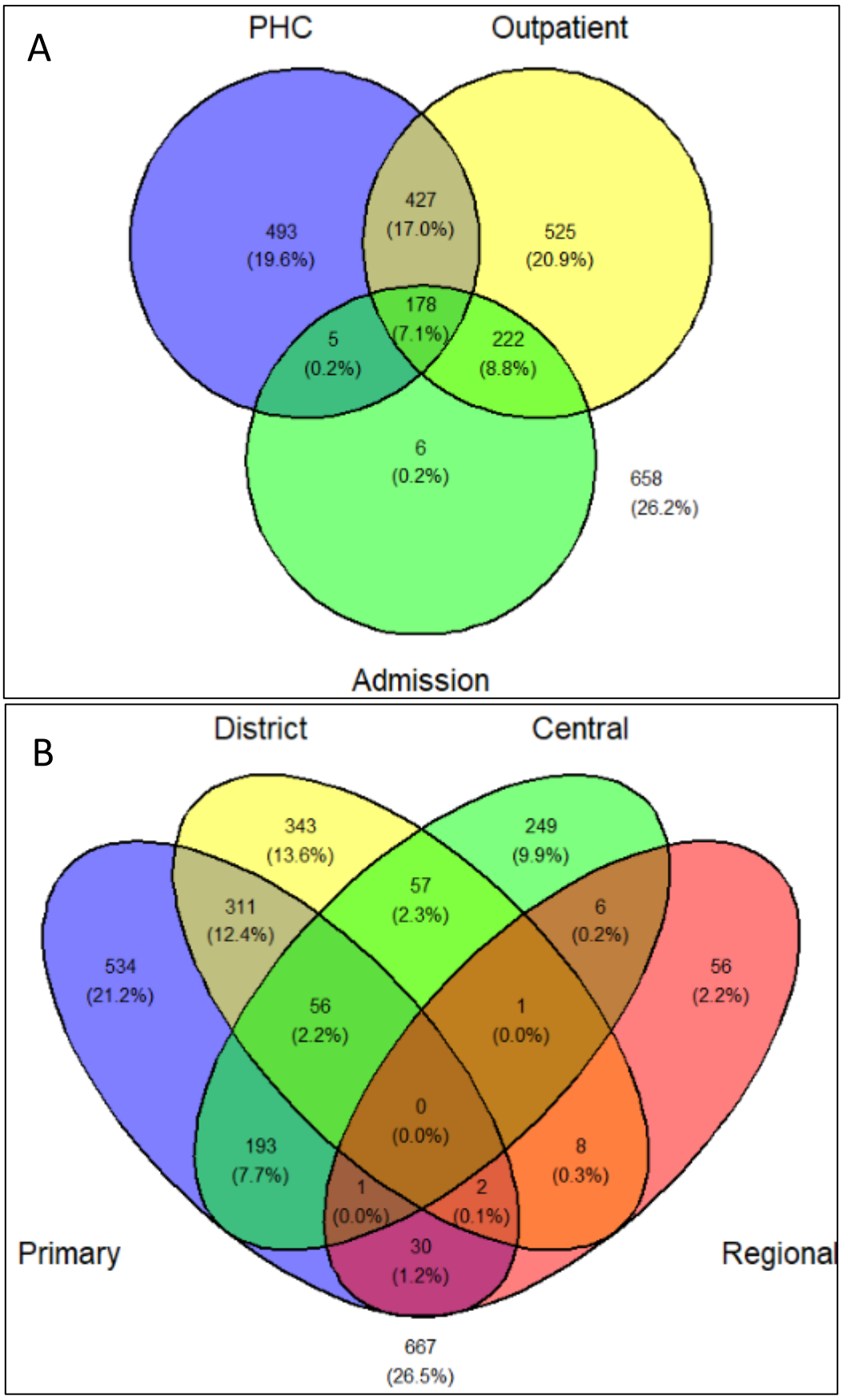


Figure 6. 7 Healthcare visit and facility types in the 30 days prior to hospital-acquired acute-on-chronic kidney disease

Intersecting circles are shown for (A) the number of primary healthcare (PHC), hospital outpatient and hospital admission visits and (B) primary healthcare, district hospital, regional hospital and central hospital facilities healthcare clients engaged with in the 30 days prior to their index community-acquired acute-on-chronic kidney disease episode. The number of the remaining individuals with no evidence of any healthcare engagement prior to their index HA-AKI episode are shown outside of the diagrams. The index admission of the HA-AKI was excluded as prior encounters were of interest. Drawn using the R ggvenn package. Abbreviations: **HA-AKI**, hospital-acquired acute kidney injury; **PHC**, primary healthcare.

Vital and kidney outcomes: Recovery after CA-AKI was complete in 13.4%, partial 9.6%, non-recovery 25.0% and uncaptured in 52.1% of individuals. For individuals with HA-AKI, there was greater complete (19.4%) and partial (15.1%) recovery. Otherwise, 32.7% had non-recovery by 7 days and 32.9% had no further SCr testing after the start of the HA-AKI episode. The mean age at in-facility death was 68.8 (SD 13.9) years for hospitalised CA-AKI and 70.9 (SD 12.4) years for HA-AKI. Figure 6. 8 and Figure 6. 9 shows the in-facility mortality following the AKI episode stratified by age category and recovery status, respectively, which in part explains the substantial proportion of episodes with non- and uncaptured recovery.

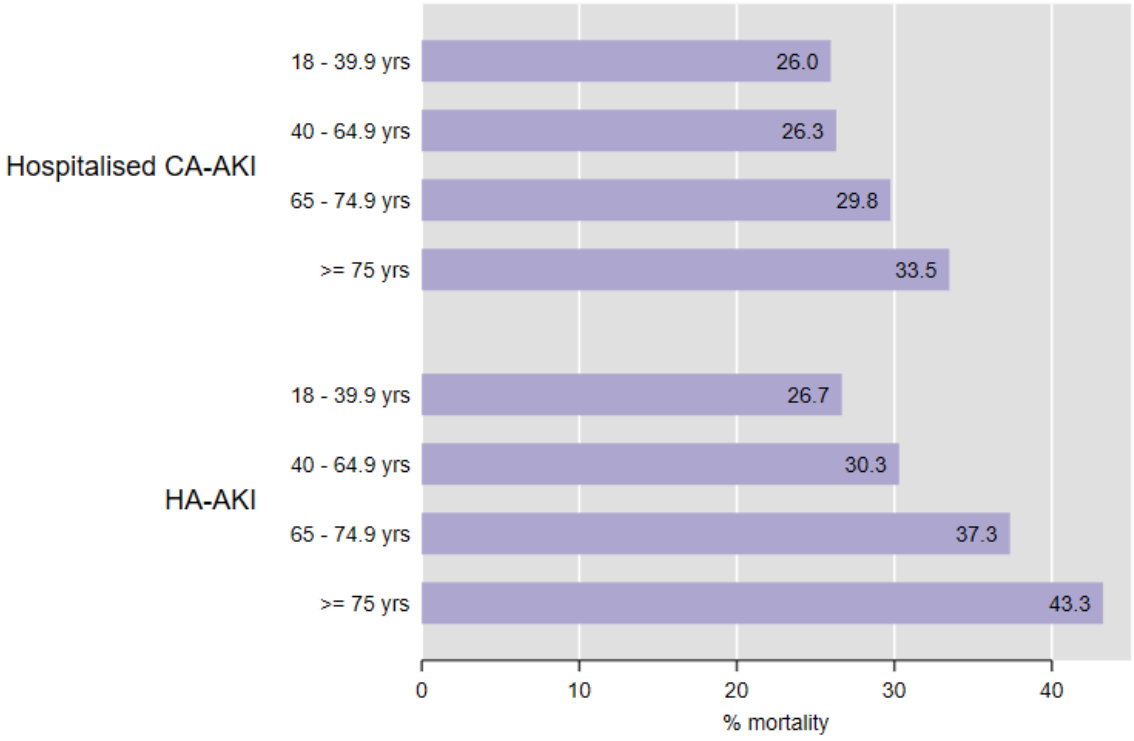


Figure 6. 8 In-facility mortality during the index admission with hospitalised CA-AKI or HA-AKI stratified by age category

The proportion of healthcare clients with hospitalised CA-AKI or HA-AKI increased with increasing age. Abbreviations: **CA-AKI**, community-acquired acute kidney injury; **HA-AKI**, hospital-acquired acute kidney injury.

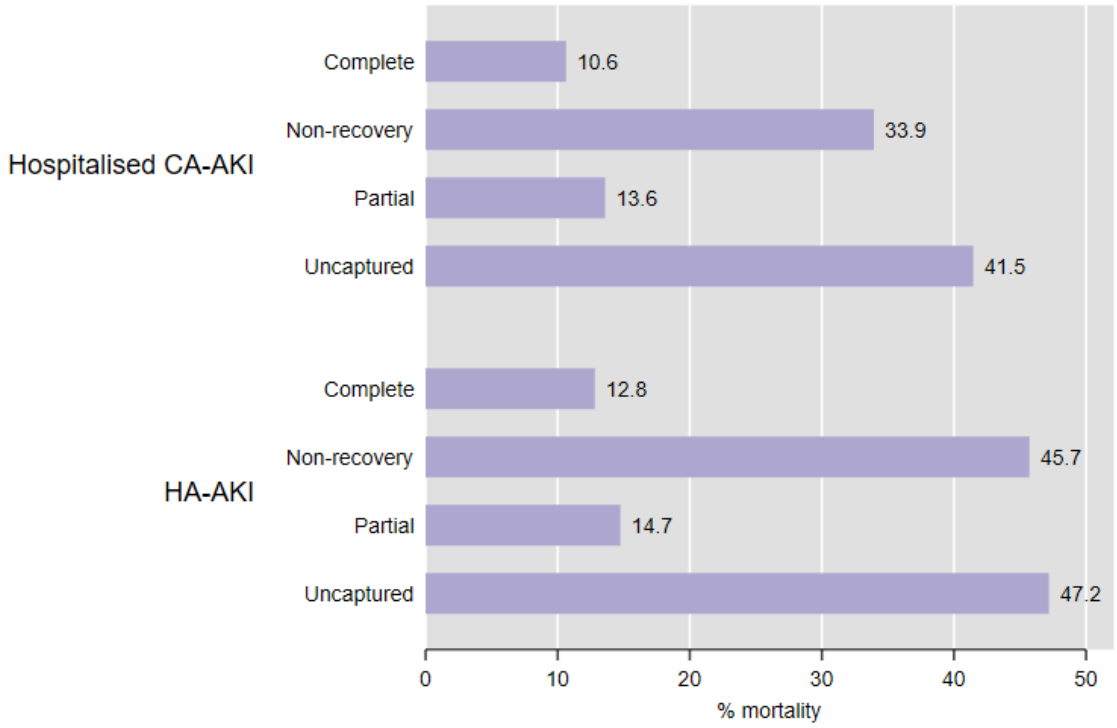


Figure 6. 9 In-facility mortality during the index admission with hospitalised CA-AKI or HA-AKI stratified by recovery status

The highest proportions of healthcare clients who died with acute-on-chronic kidney disease were those who had non-recovery of AKI by the end of the AKI episode and those who had no further SCr results following the initiation of the AKI episode. Death may therefore explain the high proportion of individuals who had an unrecorded SCr following the start of the AKI episode. Note that recovery was assessed at the last SCr result within the 7 days following the start of the AKI episode. Death may have occurred within those 7 days or later in the index admission. Abbreviations: **CA-AKI**, community-acquired acute kidney injury; **HA-AKI**, hospital-acquired acute kidney injury.

6.4 DISCUSSION

Routinely collected data by the PHDC was used to describe the incidence and traits of healthcare clients with A-on-CKD in the City of Cape Town during 2017 – 2021. CKD is an important risk factor for AKI and in this analysis, the frequency of A-on-CKD was common, often severe, and short-term kidney function recovery was often either limited or undetermined, mostly explained by high in-facility mortality.

6.4.1 Results in context

Previous literature has established that CKD is an important risk factor for AKI and A-on-CKD accelerates CKD progression and reduces the time to KRT.(213,215,216) In the current analysis, AKI was more common in those with an eGFR < 30 ml/min/1.73m^2 at CKD ascertainment. AKI incidence and poor outcomes have previously been found to increase exponentially as the eGFR decreased.(217,218) Those who experienced AKI were younger, as is consistent with AKI in people without CKD (section 5.3.1) except they had substantially more comorbid diabetes and hypertension.

The French CKD-Renal Epidemiology and Information Network (CKD-REIN) was a prospective cohort study of referred patients with an eGFR < 60ml/min/1.73m^2.(219) Hamroun et al. subsequently retrospectively identified A-on-CKD by chart review using CKD-REIN.(216) In that study, 53% of AKI was community-acquired. The study was restricted to consenting participants known to nephrology centres receiving specific kidney care, which may not be representative of all patients with CKD being seen in routine care such as those who are captured by the PHDC.(219) They found that male sex was a risk factor for AKI. In contrast to the observation that those with ascertained CKD by the PHDC were mostly female (section

5.3.1) the sex distribution was no different between those who experienced and did not experience super-imposed AKI in this analysis. However, the incidence rate of A-on-CKD was higher in male healthcare clients. Although not explored as it would be difficult to deduce the indication for SCr testing, the incidence difference might be explained by non-biological (sex) differences because of how women and men access healthcare in South Africa and are thus captured by the PHDC. The average age of those who did and did not develop A-on-CKD was above 60 years old, at which age hormonal effects on AKI development are likely to be negligible. This is supported by translational evidence of decreased ischaemia-reperfusion renal injury in ovariectomised rats that received oestradiol.(220)

While labour-intensive, CKD progression could be nephrologist-adjudicated by the previously cited authors it was not possible in this study using automated AKI detection algorithm. Nevertheless, CKD progression has previously been found to mostly be attributed to AKI non-recovery.(221) In this analysis, short-term non-recovery of kidney function was frequent and the SCr post-AKI was much higher than the AKI triggering baseline value suggesting substantial CKD progression post-AKI. It is therefore not surprising that super-imposed AKI accelerates the time to kidney failure (KF) and increases the need for KRT, and AKI prevention and its treatment should be equally aggressive, as demonstrated by others.(216,222)

It was observed that complete and partial short-term recovery was limited for both CA-AKI and HA-AKI. AKI may well have been persistent beyond 7 days into the AKD continuum. Although still invaluable prognostic information, recovery should ideally be evaluated further into the future. However, there may be further injury or insults between the index episode and future recovery assessment, especially if recovery is assessed after many weeks or

months, conflating the outcome of the current AKI. Either way, Hsu et al. have previously demonstrated poor outcomes following hospitalised A-on-CKD in Kaiser Permanente, northern California.(222) They found that of those who developed super-imposed AKI, 26% died during the index hospitalisation, which was lower than the mortality observed in the current analysis, at least for older individuals with hospitalised CA-AKI and all HA-AKI. In those with a baseline eGFR 30 – 44 ml/min/1.73m^2 and 15 – 29 ml/min/1.73m^2, 42% and 63%, respectively, developed KF within 30 days in the cited study. Also, those who experienced AKI had a 30% higher risk of death and KF. In addition, mortality was much higher than the 1-year mortality reported in CKD-REIN (12.7%). This was despite a similar AKI severity and CKD-REIN participants being on average 10 years older than healthcare clients in this analysis. The most likely explanations are that participants in CKD-REIN were in nephrology care and so may have received prompt kidney care including dialysis, and healthcare clients in the City of Cape Town are subjected to socioeconomic factors and poorly controlled comorbidities that are likely to contribute to poor outcomes.(223,224)

In the Salford Kidney Study, using a competing risk model, it was found that additional kidney episodes after the first were associated with a higher hazard of KRT initiation but not death.(213) Individuals with evidence of KRT were excluded from the analysis of A-on-CKD as the start of KRT was unknown and KRT would conflate the changes in SCr used to define AKI. It would have been pertinent, if the KRT start date were known, to investigate the effect of superimposed-AKI on the time to KRT initiation as a proxy of CKD progression. On the other hand, but not analysed, A-on-CKD may have necessitated the initiation of long-term KRT. In addition, acute dialysis was not captured, and it was not possible to confirm if those with

partial or complete recovery showed improvements in SCr because of the effects of acute dialysis.

6.4.2 Strengths and limitations

Even though the PHDC algorithm does not permit the detection of AKI once CKD has been ascertained, the detection of A-on-CKD was made possible by executing a modified AKI detection algorithm in the SCr results of persons with PHDC-ascertained CKD, overcoming the limitation of the current PHDC algorithm.

Misclassification: It was noted after manual inspection of SCr/ eGFR results that recurrent AKI was misclassified as CKD (section 5.3.1). The frequent testing of eGFR in the year before CKD was ascertained was also suggestive of acute illness. This is a disadvantage of using routinely collected data when the indication and circumstances around testing are unknown and must otherwise be inferred.(211) Because a large proportion of AKI is community-acquired, only using eGFR values generated at the time of primary healthcare visits that are more likely to represent stable kidney function compared to eGFR samples taken at the hospital, would not adequately avoid this misclassification. It is worth investigating whether current KDIGO AKI criteria accurately discern acute changes in SCr correctly in people with CKD and whether or not KDIGO severity stages are associated with the same risk of death and kidney function decline compared to persons without pre-existing CKD. Larsen et al. found that adjusting the baseline SCr for sex and time between SCr measurements decreased false-positive A-on-CKD detection.(225) Also, Xu et al. have proposed that A-on-CKD be defined by increases in SCr ≥25% (rather than the currently used 50%) based on a reference change value that was derived using intra-individual biological and analytical variation in SCr measurement.(226)

That rule detected 40% more AKI than the established 50%-rise criterion and was associated with higher mortality and CKD progression in their analyses foretelling possible improved prognostication.

Following inspection of the period before AKI was triggered, it could be seen that there was SCr 'creep'. This inflates the baseline SCr and made it more difficult for AKI to be triggered using current definitions of AKI. The algorithm used to detect A-on-CKD in this analysis fixed the baseline for the duration of the episode as an attempt to prevent this at least once AKI was already established. Of interest, a rolling 48-hour window as opposed to a static average baseline SCr resulted in superior discrimination of the risk of inpatient mortality and acute dialysis need in a recent analysis.(227) Furthermore, in the current work, the baseline SCr in the 2 and 7 days preceding the episode was on average higher than the 365-day window indicating that the SCr was probably already abnormal in the 2 to 7 days preceding the detection of AKI.

It is recognised that SCr test results may be falsely low in the post-hospitalisation period due to haemodilution, muscle loss and poor nutrition and so falsely decrease the baseline SCr used to trigger subsequent AKI.

Persistent injury beyond 7 days is possible, namely AKD, and was also considered as an extension of the AKI algorithm.(10) A conceptual framework for its inclusion as an automated algorithm was developed but was unsuccessfully implemented concurrently with the current AKI detection code (Figure 6. 10).

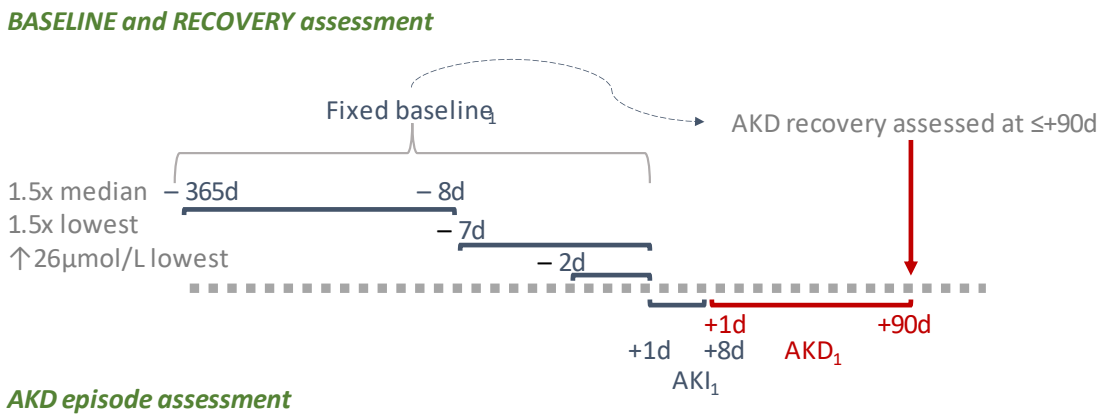


Figure 6. 10 Adaptation of the modified NHSE algorithm to include assessment of AKD detection

The dotted line represents longitudinal SCr values. On day 8 following the start of an AKI episode, AKD assessment takes place. The fixed baseline SCr, utilised during the preceding AKI episode, is reused to define new AKD. New AKD would be triggered by an index SCr on day 8 (after AKI) that was ≥ 50% of the fixed baseline – which would initiate day 0 of the AKD episode. Recovery assessment would take place on or before the end of the 90-day AKD episode. Abbreviations: **AKI**, acute kidney injury; **AKD**, acute kidney disease; **SCr**, serum creatinine.

6.5 Conclusion

For the first time and despite limitations in the PHDC AKI detection algorithm, the number of incident cases and characteristics of A-on-CKD in the City of Cape Town, South Africa, have been described. A-on-CKD was common, mostly community-acquired and short-term recovery was poor or uncaptured. Difficulty translating current consensus criteria into automated processes, uncertainty about which healthcare clients with CKD were still at risk of AKI and interconnected syndromes prone to misclassification were a challenge. The profile and frequency of A-on-CKD should be interpreted considering such.

Progression of CKD, often in response to acute insults, will next be considered in terms of the dysregulated biological mechanisms underpinning kidney function decline.

7 METHODS USED TO INVESTIGATE THE ASSOCIATION OF CARDIOMETABOLIC PROTEINS WITH KIDNEY FUNCTION DECLINE IN OLDER ADULTS WITH ADVANCED CHRONIC KIDNEY DISEASE

7.1 INTRODUCTION

In chapter one, I explained the multifactorial nature of CKD progression that involves various biological processes, including possible kidney-heart crosstalk. This chapter outlines the methods that were used to investigate the aetiological association between cardiovascular disease-related proteins, described in chapter one section 1.6.9, and eGFR decline using data from the EQUAL study. The aim was to propose some potential mechanisms of CKD progression and explain how these mechanisms might cause eGFR to decline . This analysis consists of (1) individual protein associations and (2) biological pathway associations. Reasons for the chosen model and specifications that account for confounding factors, missing data and informative censoring are detailed.

7.2 STUDY POPULATION

This analysis was undertaken using data from people enrolled in the EQUAL Study, introduced in chapter one section 1.6.8, who additionally consented to biobanking and analysis of blood samples (80% of the total cohort). Proteomic analysis by Olink® Uppsala, Sweden was performed in batches. Samples from participants in Germany (DE), the UK and Poland (PL)

were analysed together. Samples from Swedish (SE) and Italian (IT) participants were subsequently analysed separately as additional funding became available and sample processing was completed.

Samples collected in the Netherlands had not been processed at the time of this analysis and are therefore not included. Two sub-cohorts of participants were created: DE, UK, PL formed the discovery cohort and SE formed the validation cohort. The cohort was split in this way to internally validate (within the EQUAL study) that the association between protein and eGFR decline were sustained. K-fold cross-validation techniques were not possible because only results from DE, UK and PL were comparable because of the way the batches were analysed.(228) Samples from Italian participants could not be included in either cohort because laboratory plates were not bridged[18] between batches and were therefore not combinable with each other. Although possible, a separate validation was not undertaken using the Italian samples as the central biobank expressed concerns of sample mishandling.

7.3 INDIVIDUAL PROTEINS AND PATHWAYS AS THE EXPOSURE OF INTEREST

The exposures of interest were the levels of expression of the cardiometabolic proteins. Only samples collected at the baseline visit were included as there were few samples collected at KRT initiation. As detailed in section 1.6.9, two proteomic panels, reported as Normalised eXpression (NPX) units[19], were analysed on these baseline plasma samples. A complete

[18] Samples from one batch are included in the analysis of another batch to ensure that all batches analysed within a project have the same reference point. The reference point for each batch is used to standardize the protein levels and define a batch-specific level of detection value.

[19] Although the NPX unit on the $\log_2$ scale has no inherent clinical meaning, a one-unit increase can be interpreted as a doubling in protein concentration.

protein list is available in Appendix 2. Details on the panels and assay technology are provided in section 1.6.9.

Two analyses were performed: 1) each individual protein was included as the exposure, and 2) proteins were grouped together into modules representing biological pathways, detailed in the next section.

7.3.1 Bioinformatic analysis

As demonstrated in section 1.6.6, although it is convenient to evaluate proteins singly, they, in fact, act in coordinated systems. A secondary aim was to discover biological pathways that are associated with CKD progression. This was accomplished by firstly establishing which individual proteins were most highly associated with eGFR decline. Then, the biological pathways that these proteins represented were found. A pathway score was calculated, and this score was modelled in association with eGFR slope.

Biological pathways: Protein panels were carefully selected by Olink® because of the current evidence that they are empirically involved in cardiovascular disease. The proteins included in the Olink® panels analysed, therefore, have a higher chance of representing a particular biological pathway already associated with cardiovascular disease than by chance alone. For example, nine proteins are of the interleukin family and many others are immune system related potentially meaning that the proteins analysed are enriched for immunological signalling. The degree to which proteins (or genes) represent a biological pathway more than by chance (especially where multiple diverse genes/ proteins are tested) may be formally assessed using pathway over-representation analysis (also known as gene set enrichment when the interest is in differentially expressed proteins/genes).(229,230)

Reactome is an open-source, user-friendly, web-based bioinformatic database of pathways that is updated regularly and manually expert-curated.(231,232) As of version 83, December 2022, 11,350 protein isoforms and 2,601 pathways are characterised in the Reactome database. Biological pathways are organised using hierarchical GO terms, defined in section 1.7.7. Pathway enrichment analysis was performed using the Reactome knowledgebase to analyse the individual differentially expressed proteins with a protein-eGFR slope *P*-value < 0.05. The *homo sapiens* catalogue was queried using UniProt (Universal Protein) accession numbers. The analysed Olink® proteins were compared to the reference proteins catalogued by Reactome, and a Fischer-exact test was performed subject to the current canonical experimental evidence, as curated by Reactome, that a protein does or does not feature in a particular biological pathway and does or does not feature in the list of Olink® proteins. A false discovery rate *P*-value < 0.05 (see section 7.6 below) was considered as evidence that the analysed proteins were not over-represented in biological pathways more than by chance alone.

Principal component analysis: Once biological pathways were identified as described above, principal component analysis (PCA) was conducted on the individual proteins characterised by that pathway to generate a single score that best represents a complex coordinated pathway.(233) PCA summarises many correlated data points, in this case protein expression values, into linear combinations that explain the most variance[20] – called the principal components (PC) or dimensions – which are uncorrelated.(234) Conceptually, there are as

[20] Variance is a measure of dispersion calculated as standard deviation^{squared}. It is possible to compare all proteins' values' spread in this analysis because all proteins were quantified in the same NPX units of measurement and are normalised to median-zero.

many PCs as individual proteins – which is impossible to display on more than a few axes of a graph. PCA, however, transforms multi-dimensional data into much fewer PCs while still retaining maximal information. Scree plots, which order the variance explained by each PC, were used to visually decide the number of PCs that explain the most variance to keep for each pathway.(235) The retained PCs were individually entered into the model as a single variable instead of individual proteins as a complementary analysis.

Protein-Protein interactions: Interacting proteins were explored using Search Tool for Retrieval of Interacting Genes/ Proteins (STRING) database, version 11.(236) STRING-db is a web-based bioinformatic platform for examining the direct and indirect interactions between proteins, evidenced by high-throughput biological experiments or (less-evidenced) predicted interactions (for example using text-mining). The *homo sapiens* database was searched using the UniProt protein accession numbers of proteins identified as having an association with eGFR decline (as searched above). Only known associations were used i.e., those with experimental data for direct protein interactions and curated databases, but not those with only predicted interactions from low-level evidence such as text-mining.

7.4 Outcome of interest

The outcome of interest was in estimating the eGFR slope by using repeated measures of pre-KRT eGFR. Slope is the change in eGFR over time. The CKD-EPI 2009 equation[21] was used to

[21] Although there is consensus to prefer the 2021 equation without race coefficient, there may be some bias using this equation in advanced CKD and the 2009 equation is still recommended for use in Europe.(211,237) Note that the MDRD equation was used to assess eligibility in the EQUAL study.

estimate GFR from baseline to end of follow up at 6-monthly intervals (or 3-month intervals when eGFR reached < 10ml/min/1.73m^2) according to the EQUAL study protocol.(5,140) Estimated GFR was log-transformed because of non-normality of eGFR (Figure 7. 1).

Box 7. 1 shows features of this longitudinal repeated measurement data requiring consideration in the analysis.

- Unequally spaced eGFRs for individuals (3/6 months, may have missed visits; routinely collected data, not at pre-specified study visits)
- The number of eGFR measurements differs between patients
- Patients are enrolled at different points in calendar time (staggered entry)
- Patients are enrolled at different baseline eGFR
- The relationship between eGFR and time is non-linear and dynamic, affected by non-GFR factors as well
- Patients may have different slopes (steep vs slow decline vs improvement vs catastrophic)
- eGFR values are more closely related within individuals than between individuals (dependency)
- eGFR values are more closely related when measured closer together than further apart in time
- eGFRs will not be 'measured' after some events
    - Early drop out may be informative (non-ignorable missing data)
    - Some events (e.g. death) may compete for other events (e.g. dialysis/ transplantation)

Box 7. 1 Features of the longitudinal EQUAL data

Figure 7. 1 illustrates the distribution of eGFR values. The reason for right-skew is notable given that the EQUAL study is of people with advanced CKD in whom the eGFR would be expected to be much lower than 60 ml/min/1.73m^2. The inclusion criteria, however, only required a single eGFR of < 20ml/min/1.73m^2 prior to study entry which meant that some patients were included with a previously spuriously low eGFR, although patients with ostensible AKI were excluded by the EQUAL study protocol.

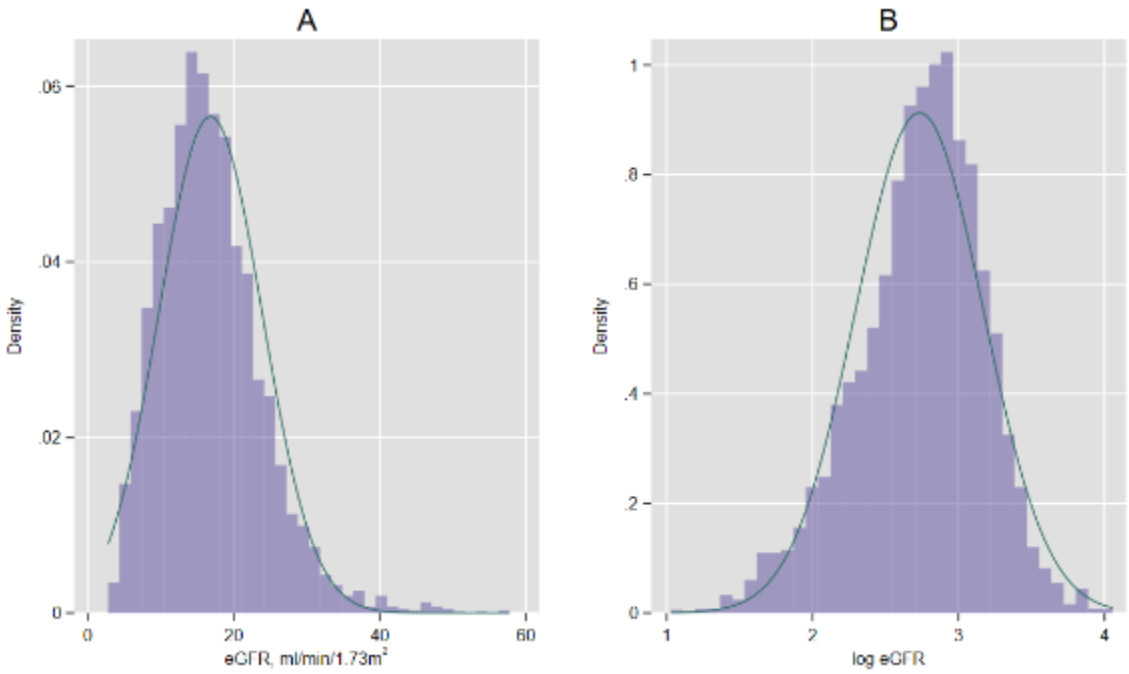


Figure 7. 1 Pre-dialysis eGFR distribution

A. Untransformed eGFR and B. log-transformed eGFR Abbreviations: **eGFR**, estimated glomerular filtration rate.

7.4.1 Primary analysis of repeated measures (of eGFR)

Individuals may have different starting eGFRs when they entered the study (intercept) and different trajectories of eGFR decline (slope). eGFR measurements[22] are clustered per person, individuals are in turn clustered per country creating a hierarchy. Multilevel (also known as hierarchical) linear mixed models (MLM) with random effects were therefore used to analyse the repeated measures of eGFR.(238) The MLM model framework allows the intercept and slope to be calculated individually, hence modelling random effects, instead of the population average i.e., fixed effects which are assumed to be unchanged across the entire study population, as depicted in Figure 7. 2. The MLM also allows for unbalanced study design, as used in the EQUAL Study, with irregular times between study visits at which data collection was performed.(140)

Generalized linear mixed model (GLMM) with Poisson family and log-link are preferred to including log(eGFR), necessary because of non-normally distributed eGFR values, in a linear mixed effect model because log(eGFR) is not readily interpretable; a GLMM was therefore used in this analysis.(239,240) The log-link rescales the relationship between slope and time to a logarithmic distribution and so negates the need for splines in the setting of a nonlinear relationship.(225) Robust standard errors were calculated because model-based standard errors may be biased when repeated measurements are not independent of each other within and between clusters and relaxes the assumption that the variance must equal the mean in Poisson regression.(242) Protein*time interaction coefficients (multiplied by 100) were interpreted as % change in eGFR per year per doubling of protein concentration. In addition,

[22] The term 'measurement' here is in the statistical sense and is not intended to mean the direct measurement of GFR.

predictions of the average slope, after adjustment for independent covariables (section 7.4.4), were visually plotted. Follow-up may have ended prematurely because KRT was initiated, the patient died, the patient was lost to follow-up or study completion had occurred (at 4 years), whichever came first.

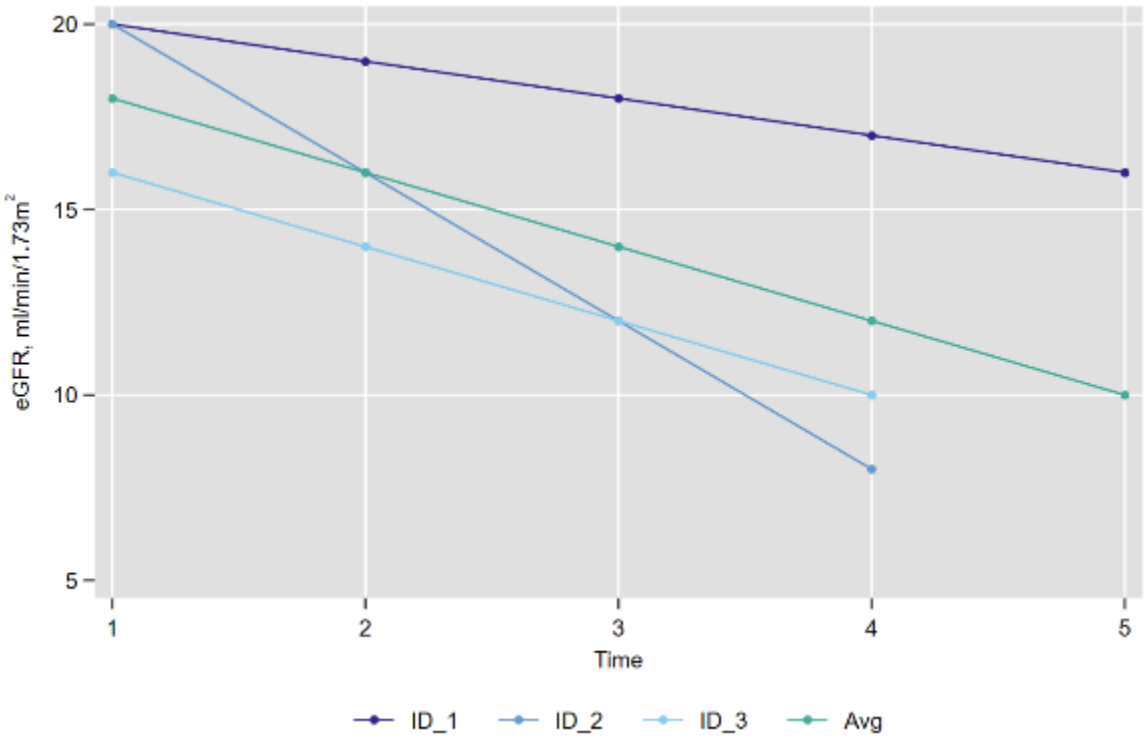


Figure 7. 2 Conceptual diagram of eGFR slope

Traditional regression analysis would calculate a single intercept and slope assuming that all individuals decline their kidney function at the same rate i.e., a population average ('**Avg**'). In reality, individuals (**ID_**) may have different starting eGFR's and slopes. Abbreviation: **eGFR**, estimated glomerular filtration rate.

7.4.2 Levels of the multilevel model

There are three hierarchical levels defined which are nested within the higher order level (Figure 7. 3):

3. **Study country** – it is to be expected that populations would differ systematically in the starting eGFR at study entry because of referral to specialist nephrology service practices. Also, the trajectory in eGFR slope, pre-dialysis care, juncture at which KRT initiation is considered, and mortality outcomes may differ between countries,
2. **Inter-individual** variability *between* study participants,
1. **Intra-individual** variability *within* the same person.

Random effects in this analysis included individual level (level 1 and 2) parameters only for model simplicity and because the main purpose of mixed effects modelling was to account for repeatedly measured eGFR. Country was instead included as a fixed effect.

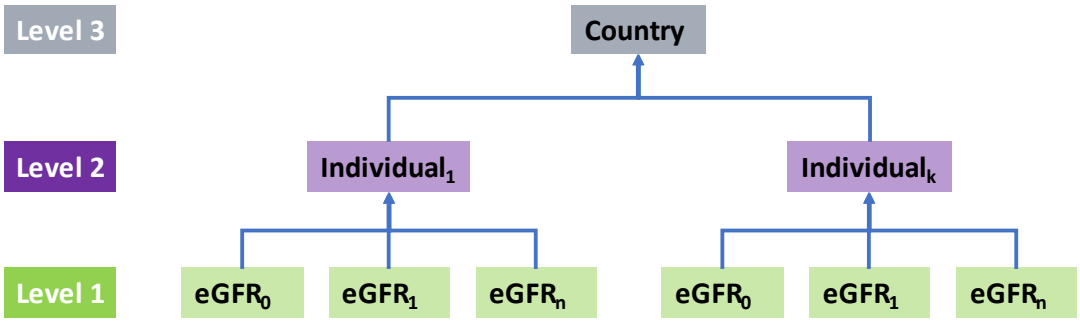


Figure 7. 3 Hierarchy of multilevel effects

eGFR measurements are measured repeatedly within the same person (level 1) and measurements differ between people (level 2). These research participants are nested within a country (level 3). Abbreviations: **eGFR**, estimated glomerular filtration rate.

7.4.3 Covariance

eGFR values within the same individual are likely to be correlated more than eGFRs between people, particularly when more frequent and closer in time. The covariance structure therefore becomes important when considering the intra- and inter-individual variation of eGFR within and between patients and should be estimated as close as possible to the true relationship. Different designs are available: independent, compound, autoregressive, and unstructured.(243) An unstructured organisation was selected for this analysis which makes no assumption about the relationship between pairs of eGFR values regardless of the time between measurements and changing fluctuations within the same person.

7.4.4 Confounding and mitigation of other biases

The causal association between expression levels of cardiometabolic proteins and eGFR slope may be affected by other measured or unmeasured factors. The assessment of the relationship may be biased if not accounted for, or conversely, if inappropriately adjusted for. Causal diagrams (directed acyclic graphs, DAGs), established on causal inference theory, were used to set out my assumptions about the relationship between the independent variables on the exposure and outcome and independent variables between each other. Inspection of the causal structure helped to appropriately account for confounding factors and exclude mediating or collider biased pathways (see Figure 7. 4).(244,245) DAGs include nodes (variables) connected by unidirectional edges (lines with arrows). When two edges converge on a variable, that variable is called a collider. Indirect 'backdoor' paths may induce an erroneous association between extraneous variables, thus producing bias. Conditioning on a collider, for example, may induce collider bias.(246) Given that there were 175 proteins i.e., exposures, a general framework was created for practical reasons, using a pragmatic approach

that took into consideration clinically important risk factors for CKD progression, influences on protein levels and study country as a proxy for geographic policies on treatment, referral to nephrology service practices and possible differences in vital and kidney outcomes. Five sequential model adjustments were conducted:

0. Unadjusted
1. Model 0 + sex, age, country
2. Model 1 + systolic blood pressure, diabetes mellitus status, and primary renal disease (as defined by the European Renal Association – European Dialysis and Transplant Association)(247,248)
3. Model 2 + albuminuria (albumin: creatinine ratio)(249)
4. Model 3 + medications (renin angiotensin aldosterone system inhibitors [angiotensin converting enzyme inhibitors, aldosterone receptor blockers and mineralocorticoid receptor antagonists] and β-blockers)(250,251)

Model 4 was considered the primary analysis but since medications would not be expected to influence all protein levels, model 3 was also considered of secondary interest. Furthermore, an assumption was made that medications confounded the relationship between protein levels and slope; comorbidities, such as hypertension, did not mediate the relationship between protein level, medications and slope (protein → hypertension → antihypertensive → slope). This was decided as although the antihypertensive was prescribed to treat hypertension, hypertension was proxied by systolic blood pressure (BP) and the antihypertensive would directly affect the level of BP. Systolic BP was measured using a standardized study protocol. Diabetes mellitus status, medications and PRD were recorded by the treating nephrologist. Only variables collected at the baseline visit were used i.e., variables

were not allowed to vary over time. This was decided because future co-variables cannot influence past values of proteins measured exclusively at the baseline visit. Country was included as there may be factors related to the study site that determines how patients were managed. Some models for the discovery analysis failed to converge for some proteins (n = 3 proteins [model 2], 13 proteins [model 3] and 12 proteins [model 4]). Country was removed as a covariable in these models to achieve model convergence. Baseline eGFR was included in the repeated measures eGFR and not adjusted separately.

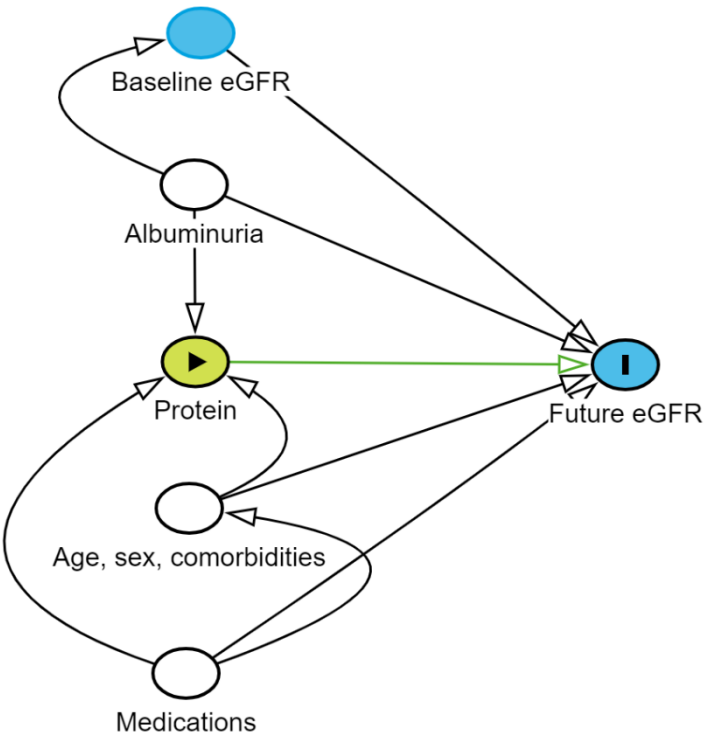


Figure 7. 4 Postulated exposure-outcome relationship in the setting of other variables

Nodes (variables) are connected by arrowed edges. The exposure(s) of interest were Olink® proteins and the outcome was eGFR slope. Blue nodes represent the only direct ancestors (causes) of the outcome, other than the exposure (green). The green edge represents the direct analysis of interest (front door). All back doors through the white nodes are blocked after adjustment. Age, sex, systolic blood pressure, diabetes status and primary renal disease are represented by the single 'comorbidities' node for simplicity, included from the baseline visit only (time invariant). Please note the hidden arrowhead from 'Medications' to 'Future eGFR'. Abbreviations: eGFR, estimated glomerular filtration rate. This DAG was produced using the DAGitty web-based environment.(252) Abbreviations: **eGFR**, estimate glomerular filtration rate.

7.5 SENSITIVITY ANALYSES

Two sensitivity analyses were conducted, as detailed below, to confirm the robustness of the primary analysis in the presence of informative censoring and substantially missing ACR data. Sensitivity analyses were conducted on both discovery and validation sub-cohorts.

7.5.1 Joint modelling

Patients could withdraw follow-up before the scheduled completion of the study because of death, loss to follow-up or KRT initiation. When this happens, patients are said to be censored (graphically depicted in Figure 7. 5). Censoring may be noninformative when these losses to follow-up are unrelated to the study or worsening kidney function. Standard methods to estimate the rate of change or slope of a longitudinal outcome assume that dropouts from the study are noninformative and this is an important model assumption. In some situations, patients with CKD might drop out of studies because of the severity of their kidney disease.(253) Patients with advanced kidney disease for example would often be expected to initiate KRT once KF develops, which was of particular concern in the EQUAL study since patients had advanced kidney disease and were expected to reach KF rapidly.(140) Alternatively, those with stable or unstable kidney function who choose conservative care may be referred back to primary care for further follow-up. Since native kidney function can no longer be quantified after the initiation of KRT or loss to follow-up, eGFR measured after this time becomes unobserved i.e., missing, and estimation of the eGFR slope may therefore be biased. The rate of decline in kidney function, among other factors, may have been considered in the decision whether to initiate KRT. Censoring in this situation is termed *informative* and special modelling methods are required to take this into account.(253–255)

Although people who initiate KRT may go on to later die with or from their kidney disease or be LTFU, the first occurrence primarily inhibits the measurement of future eGFR (data become missing).(256) These first events 'compete' with each other in that they are mutually exclusive. In some patients with CKD, especially those with advanced disease and those who are older with multi-morbidity, the risk of death is higher than initiating KRT.(257) In a survival analysis, simply censoring events in this situation is undesirable because estimates are biased when the probability of one occurrence is higher than another and those who are censored are potentially still at risk of experiencing a competing event.(258) The time to these events is therefore 'cause-specific' when the research question is aetiologic. In the cause-specific competing risks framework, people are removed from the dataset after experiencing the first event because they are no longer at risk of experiencing a competing event.[23]

The following events were taken as competing in this analysis:

1. KRT initiation
2. All-cause mortality
3. LTFU
4. If a person did not initiate KRT, was not LTFU or did not demise, they were considered alive at last follow up (administrative censoring at the time of last database update or study completion)

[23] In predictive research, the sub-distribution i.e., Fine and Gray model is used instead in which individuals remain in the dataset after the first competing event.(259)

Reasons for leaving the study may have been because the renal clinic discharged the patient back to primary care, the patient moved away from the study site, or the patient withdrew consent.

As a sensitivity analysis, joint models (JM) were used in the current study to explore the possibility and effect of informative censoring and competing risks as the eGFR slope may potentially differ between persons with different censoring events, see Figure 7. 5.(260,261)

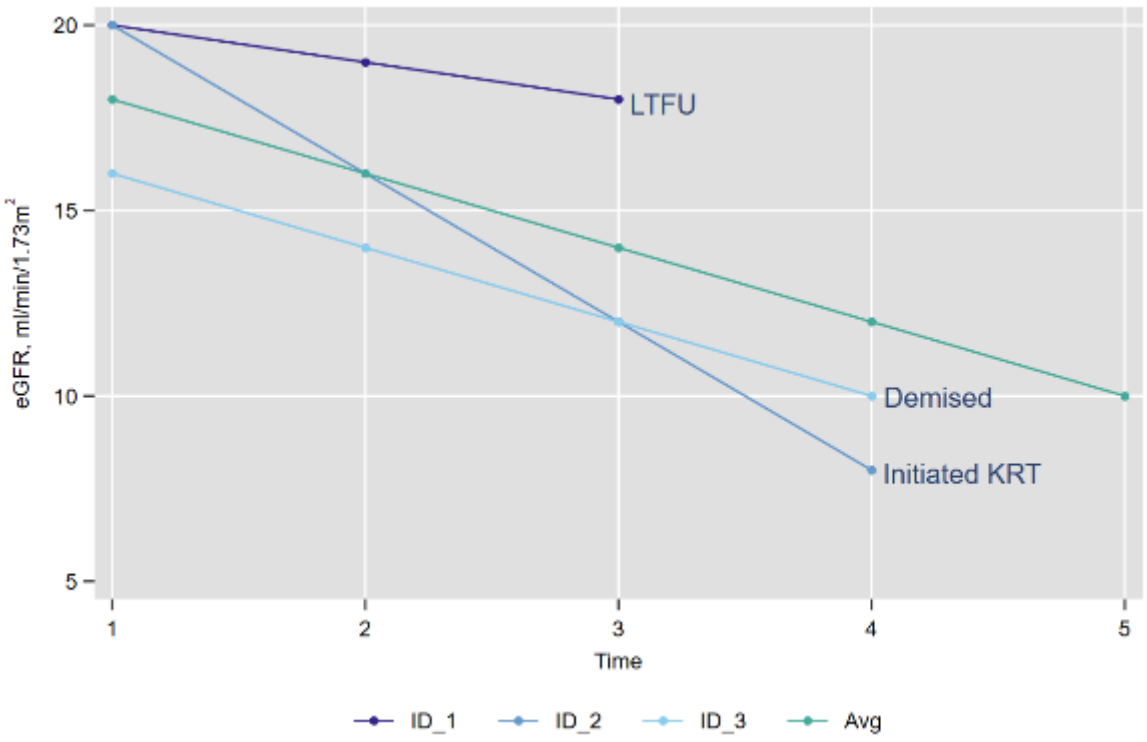


Figure 7. 5 Conceptual diagram of the end of eGFR observation after competing censoring events

Estimated GFR is unknown after patients initiate KRT, die or are lost to follow up. eGFR slope, the outcome in this analysis, becomes censored after this time. The lines represent individuals (ID_) with their own personal trajectories and censoring reasons, while 'Avg' line shows the

overall average of all individuals. Abbreviations: **eGFR**, estimated glomerular filtration rate; **LTFU**, lost to follow up; **KRT**, kidney replacement therapy.

The joint model framework includes two or more sub-models that are analysed simultaneously:

1. MLM longitudinal model (described already in section 7.4.1) and,
2. Time-to-event (aka survival) model(s) which will be described below.

Parameters computed by the longitudinal model are incorporated into the estimation of the survival model and vice versa. The advantages of using JMs are to allow for the effects of potential informative censoring in the estimation of the longitudinal model by stipulating how the time to early dropout (by KRT initiation, death and LTFU) affects eGFR slope. Only the longitudinal model parameters i.e., the slope coefficient, were of interest and reported. Two user-written Stata commands were used to execute the JM: `stjm`, which does not allow competing risks, and the more flexible `merlin` v2.1.5 (*Mixed Effects Regression for LInear, Non-linear and user-defined models*), which allows more than one MLM and survival (for the purposes of competing risks) models and splines, both written by Michael J Crowther.(262,263) `merlin` incorporates multiple *components* (the sub-models) which consists of various *elements* which build the specification of the model. Elements include the independent and dependent variables, nonlinear terms, interactions, random effects, and distribution families.

Longitudinal sub-model: A random intercept and slope were specified. Iterations of the longitudinal model sequentially adjusted included models 0 – 4.

Time-to-event sub-model: The time from which eGFR measurements were recorded in the study i.e., the baseline visit, was taken to be when the study participant was first at risk in all time-to-event analyses and the timescale was measured in years until the first censoring event (either KRT initiation, death, LTFU or otherwise censored as alive at last follow-up).

Each competing risk component sub-model was included separately. In other words, in addition to the longitudinal MLM model, a sub-model for each time to KRT, death, LTFU and administrative censoring was specified. Administrative censoring occurred when end of follow-up was reached, or data extraction was undertaken and no occurrence of KRT initiation, death or LTFU had otherwise been experienced. The KRT time-to-event sub-model was adjusted for age, sex, country, and primary renal disease. The death time-to-event sub-model was adjusted for age, sex, diabetes mellitus status, systolic blood pressure, country, primary renal disease and ACR.

7.5.2 Missing data

There were some missing data for Olink® proteins. Superoxide dismutase 1 protein failed quality control in 97 individuals in the discovery sub-cohort. None were missing protein data in the validation sub-cohort. Age, sex and country had zero missing values. Information was missing for <1% for the co-variables diabetes mellitus status, primary renal disease and systolic blood pressure. Estimated GFR (the outcome) was missing for <1% and missingness of the repeated measure variable is inherently handled within the multilevel model. Of most concern, baseline ACR was missing in 43%. Data were collected from patient questionnaires at the time of routine clinic visits.(140) Biochemical results were recorded from medical

records at these associated visits and testing was not specifically requested for the purpose of the EQUAL study. It would be expected that every non-anuric[24] patient visiting the kidney clinic would undergo a urine test (at the very least a urine dipstick) and no one should be expected to be anuric at the baseline visit with an eGFR of 20 ml/min/1.73m^2 and so be unable to produce urine. Proteinuria would therefore be readily detected and a urine ACR or PCR requested, and its result recorded. It was hypothesised that significant proteinuria would be more likely to be detected and recorded and so its missingness was dependent on unobserved proteinuria levels.(264) Alternatively, urine collection or result recording may have differed because of site-specific practice patterns.

Handling of missing data: The simplest solution to missing data, and the default of most statistical models and software, is list-wise deletion resulting in a complete case analysis of individuals with available data across all variables (exposure, independent variables and outcome). This may be pragmatic but potentially problematic for two reasons. Firstly, excluding observations or individuals missing some data (even one value for a particular variable) leads to a loss of sample size and decreased precision in model estimates. Secondly, this approach assumes that the resulting dataset is a true representation of the whole study population with observations excluded. This may not be true if the individuals with incomplete data are different to those with complete data leading to biased model estimates. If the actual missing values or reasons for missingness can be recovered, then these problems may be overcome. For clinical data such as diabetes status, other related data – the use of antidiabetic medication, mention of the type of diabetes – were used to assume that diabetes in fact,

[24] Urine production < 50ml/day

coexisted. Missing albuminuria data were recovered by converting measured protein: creatinine (PCR) to albumin: creatinine ratios.(265) Actual ACR and converted PCR to ACR values for those with both ACR and PCR available were mostly comparable; outliers were probably incorrectly recorded in another unit of measurement or a decimal place was erroneously added (Figure 7. 6).

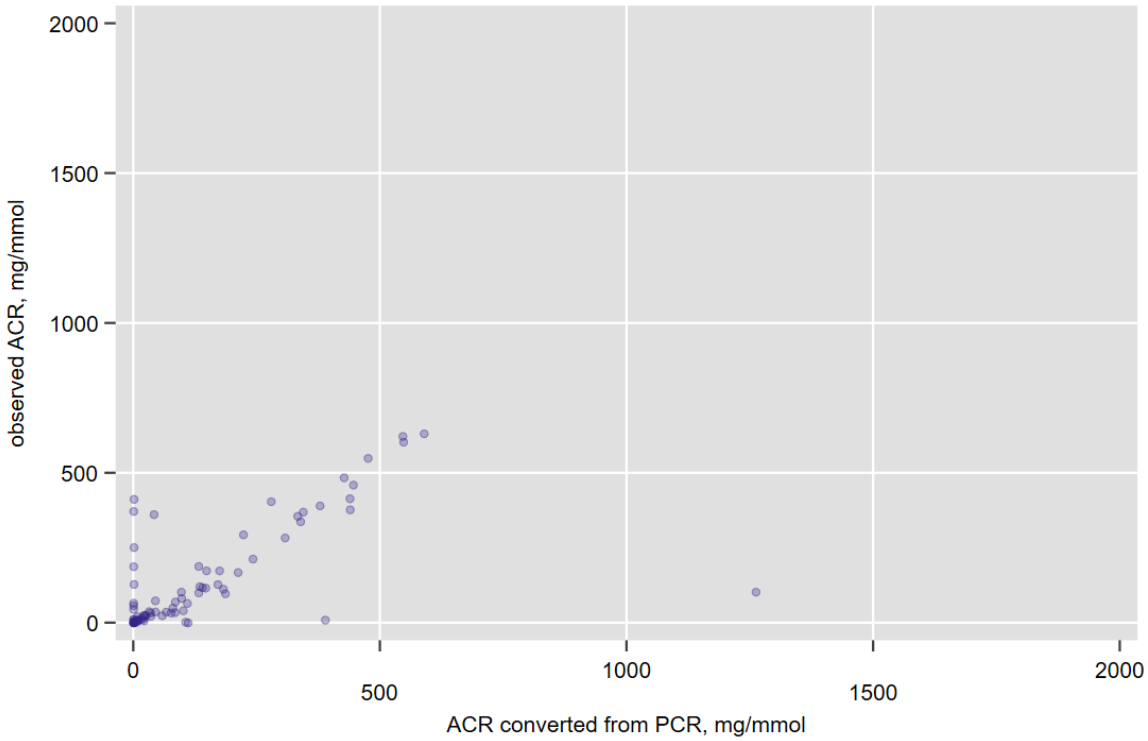


Figure 7. 6 Comparison between actual and converted ACR

The actual ACR values are compared to the predicted urinary ACR values using Weaver et al's equation to convert PCR to ACR. There may have been measurement error due to incorrect recording of PCR and ACR in cases where the ACR or converted ACR measure was close to zero but the corresponding ACR or converted ACR was much higher (follow data points along 0 mg/mmol). Abbreviations: **ACR**, albumin: creatinine ratio; **PCR**, protein: creatinine ratio.

Missingness mechanisms: Rubin (266) proposed three mechanisms of missingness that explains systematic differences in the observed and missing data:

1. **Missing completely at random (MCAR)**: there are no systematic differences between the observed and unobserved data,
2. **Missing at random (MAR)**: any systematic differences between the observed and unobserved values can be fully explained by data that were observed,
3. **Missing not at random (MNAR)**: associations with the observed data cannot explain all systematic differences between the observed and missing data. (267)

Systematic differences may be explained by observed data in the exposure of interest, the outcome of interest or other data. The missingness mechanism therefore further depends on whether the missingness can be explained by the exposure, outcome, or independent model variables. In a linear regression, a CCA is deemed biased when the chance of a complete case i.e., ACR is not missing, is dependent on the outcome alone or in combination with the exposure and/ or confounders.(267) Following adjustment for age, sex, systolic blood pressure, diabetes status, study country, primary renal disease, RAAS-inhibiting drugs and β-blockers, a logistic regression model was used to assess the odds that the eGFR slope (outcome) would have on someone having ACR observed or not. Based on these results (presented in section 8.1.5) and a DAG drawn (Figure 7. 7) showing the assumed relationship between the exposure, outcome and included independent variables in the model, a CCA was thought to be unbiased and so was conducted as the primary analysis.

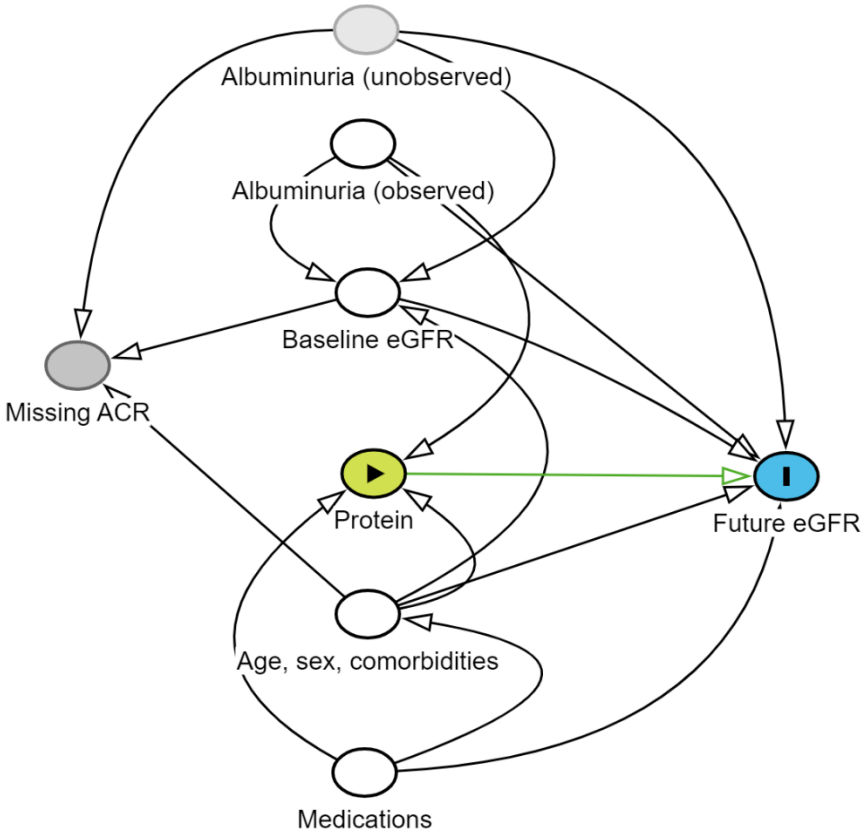


Figure 7. 7 Potential missing data mechanism for albuminuria

In this DAG, the assumed relationship between the exposure (green), adjusted-for independent variables (white), outcome (blue) and unobserved variables (grey) are shown. There is no relationship between the outcome and chance of having a missing ACR. Note the hidden arrow from 'Medications' to 'Future eGFR'. Abbreviations: **ACR**, albumin to creatinine ratio; **eGFR**, estimated glomerular filtration rate.

Multiple imputation: Multiple imputation (MI) is most appropriate when a CCA is deemed to be biased and recovery of missing data may lead to improvements in precision and efficiency.(268) As a sensitivity analysis, MI was conducted to compare the slope estimates with the primary analysis assuming a MAR missingness mechanism. MI uses a model to randomly sample values of the missing data ('imputed values') from their predicted distribution based on the observed data. The analysis of interest is then computed using each separate imputed dataset and the estimates and their standard errors are then combined using Rubin's rules.(269) This is preferred as MI also considers the uncertainty in deriving the estimates. MI assumes that data are MAR (or MCAR) because any systematic differences between complete and incomplete cases can be explained by observed values (of the missing variable). When the missingness mechanism is thought not to be MAR, 'auxiliary data', information other than the variables included in the main analysis, may be used to explain its missingness thereby producing a conditional MAR mechanism. That is, a MNAR mechanism is transformed into an MAR mechanism. There were unfortunately no auxiliary variables, other than those already included in the substantive model.

Ten donor draws predictive mean matching (PMM) were used to impute missing values of ACR with 50 multiple imputation datasets (approximately one imputed dataset for each 1% missing ACR).(270) PMM replaces the missing value with a random observed value in the nearest vicinity of the imputed prediction as recommended for the imputation of non-normally distributed data.(271) All variables in the substantive model (the analysis of interest), including the exposure, independent variables, outcome (repeated measure eGFR) and the protein*time interaction, were included in the imputation model. Rubin's rules were used to

combine the imputation model estimates.(269) The Stata command `mi` was used to perform the multiple imputation procedures.

A specific 'not at random fully conditional specification' is available to impute MNAR data but was not undertaken in this analysis.(272) Although special considerations are required in the context of multi-level modelling, single-level MI was used because only baseline ACR was used in this analysis and so did not vary over time within the same patient.(273,274)

7.6 MULTIPLE TESTING

There is a high probability of observing a spurious relationship purely by random chance between protein and eGFR slope because multiple (175 proteins) hypothesis tests were performed.(275) An adjustment for multiple testing was required; by one of two methods, either false discovery rate (FDR) or family-wise error rate (FWER). The FDR is the number of falsely rejected null associations that would be expected by chance given the greater the number of proteins (in this case) that are tested while FWER is the probability of making *at least* one type I error[25]. The Bonferroni correction is the most widely used FWER procedure, however has been criticised for excessively penalising the probability of declaring at least one test falsely significant which renders the technique overly conservative.(276) A less stringent approach in terms of detecting true positives is the Benjamini-Hochberg (B-H) FDR procedure, conducted in the following steps, subsequent to computation of the model parameters(277):

[25] Type I errors represent false positive conclusions whereas type II errors characterize false negative conclusions.

1. *P*-values for each protein were sorted in ascending order,
2. the critical B-H value = (*P*-value rank/ the number of proteins)*0.05[26] was calculated,
3. the unadjusted *P*-values were compared with the critical B-H value; those with the largest *P*-value < the critical B-H value were identified,
4. all proteins above this level were considered significant, inclusive of the current level.

The user-written Stata command `multproc` was used to automate the B-H procedure. (278) Proteins meeting $P_{FDR} < 0.05$ in the Discovery sub-cohort were taken to validation. Proteins with $P < 0.05$ were identified as proteins that successfully validated.(102) Unfortunately, it is not possible to estimate 95% confidence intervals using the Benjamini-Hochberg procedure.

7.7 SAMPLE HANDLING AND STORAGE

Plasma samples, collected in gel-separated tubes, were refrigerated at -80° Celsius at the local study site and subsequently shipped to a national laboratory in each country. Batches of samples were transferred to a biobank in Würzburg, Germany (for samples originating in UK, Poland, Italy, and Germany) or a biobank in Uppsala, Sweden (for samples originating in Sweden). Samples were thawed, aliquoted, plated and refrozen before being sent to the central Olink® laboratory for analysis. The candidate had no involvement in, or oversight of, any sample handling or processing.

[26] The conventional 0.05 was selected as the FDR in this analysis.

7.8 ETHICAL CONSIDERATIONS

Written informed consent was provided for all participants included in the EQUAL study to collect clinical information and, separately, bio-samples. Ethics approvals in participating countries were obtained prior to commencement of the study. NHS Health Research Authority approved the original study in the UK (13/SW/0016). De-identified data were used in this analysis.

7.9 CONCLUSION

In this chapter, I have detailed the methods used to analyse the relationship between cardiometabolic proteins and biological pathways with kidney function decline using data collected in the EQUAL study of older people with advanced kidney disease. A summary of the analysis is provided in Figure 7. 8.

I have highlighted several important considerations about these data including the use of repeated measurements of eGFR, mitigation of bias, concern for informative censoring and competing risks, handling of missing data and multiple hypothesis testing.

The results and discussion of these primary and secondary analyses are presented next.

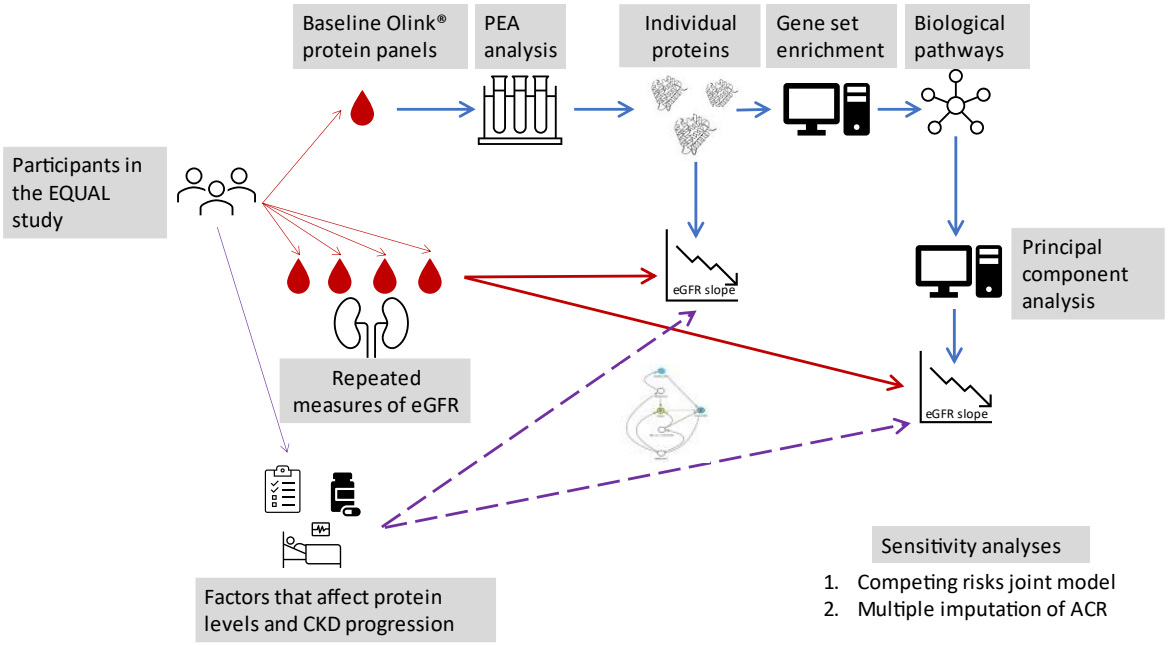


Figure 7. 8 Summary of analysis to investigate protein and biological pathway associations with eGFR slope

Blood samples and clinical information were collected from consenting participants in the EQUAL study. Baseline samples were sent for analysis for two cardiometabolic protein panels using Proteomic Extension Assay (PEA) technology. Gene set enrichment was used to identify biological pathways that these highly differentially expressed proteins mostly represented. The concentrations of the proteins that were represented in the biological pathway were combined using principal component analysis. Principal components and individual proteins were analysed in association with eGFR slope, adjusted for clinical factors, as appropriate. The individual protein-slope analysis was supplemented by two sensitivity analyses to account for informative censoring and missing ACR values. Abbreviations: **eGFR**, estimated glomerular filtration rate; **CKD**, chronic kidney disease; **PEA**, proximity extension assay; **ACR**, albumin: creatinine ratio.

8 CARDIOMETABOLIC PROTEIN AND PATHWAY ASSOCIATIONS WITH KIDNEY FUNCTION DECLINE IN OLDER ADULTS WITH ADVANCED CHRONIC KIDNEY DISEASE

8.1 RESULTS

In the previous chapter, the methods used to analyse the cardiometabolic protein-eGFR slope relationship were presented. In this chapter, the results, analysis and discussion are given. I will start with a description of the cohort, the results of the individual protein primary analysis and then the results of sensitivity analyses. In the discussion section, I discuss possible mechanisms why particular associated proteins and pathways might underlie kidney function decline.

8.1.1 Patient characteristics

Overall, a complete case analysis was undertaken on 501 individuals, approximately equally split between discovery and validation sub-cohorts (Table 8. 1). Baseline characteristics are shown in Table 8. 1. Age was comparable between sub-cohorts and slightly more females were included in the discovery sub-cohort (discovery: 41% versus validation: 28%). The discovery sub-cohort featured mostly individuals from the United Kingdom. A total of 4,472 pre-KRT eGFR values over time were available to estimate eGFR slope. Only 90 values were available after 4 years and so slopes were predicted ≤ 4 years only to avoid erroneous slope estimates because of sparse eGFR points beyond this time. The discovery cohort averaged fewer eGFR

values per person (median 4; IQR 2, 6) compared to the validation sub-cohort (median 7; IQR 3, 9) but the baseline eGFR was very similar, averaging 17.7 ml/min/1.73m^2 overall. Although an eGFR of <20ml/min/1.73m^2 in the past 6 months was mandatory for eligibility, some individuals had eGFR values close to 60ml/min/1.73m^2 at the time of inclusion (Figure 8. 1).

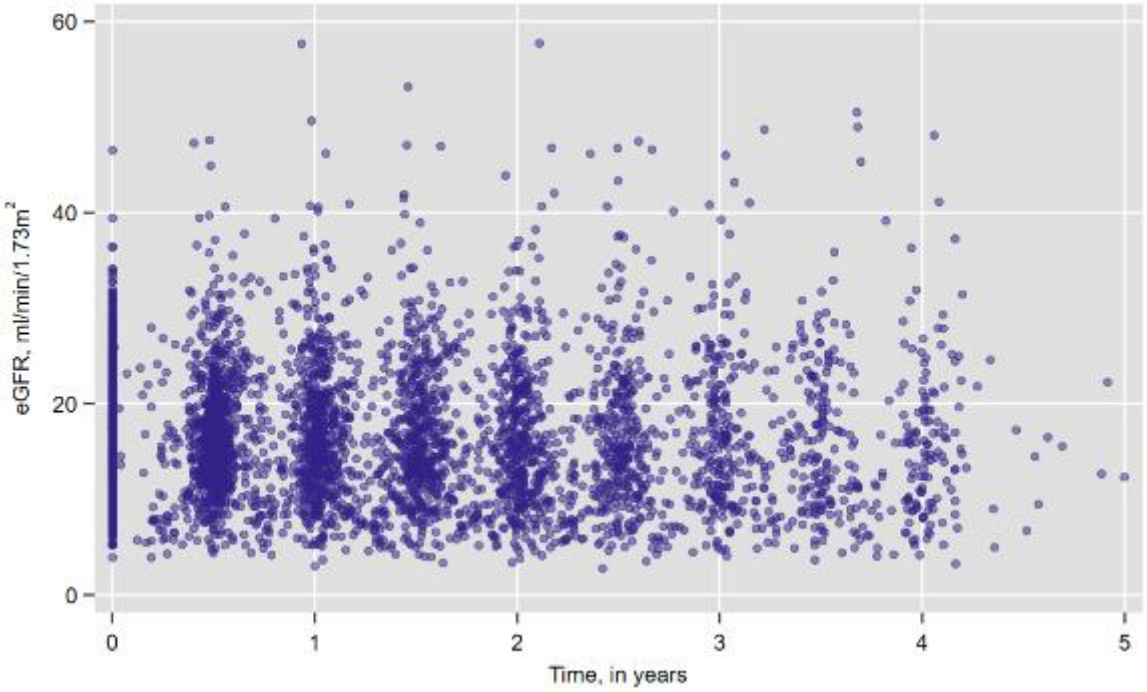


Figure 8. 1 Scatterplot of eGFR over follow-up time

eGFR was calculated using the CKD-EPI 2009 equation. Abbreviations: **eGFR**, estimated glomerular filtration rate; **CKD-EPI**, Chronic Kidney Disease Epidemiology.

	Overall cohort, N = 501	**Discovery cohort**, n = 254	**Validation cohort**, n = 247
Age in years, median (IQR)	75 (70, 81)	76 (69; 81)	75 (70; 80)
Female sex, %	34%	41%	28%
Country, %			
Germany	9.2%	18%	0%
United Kingdom	30%	59%	0%
Poland	12%	23%	0%
Sweden	49%	0%	100%
Comorbidities, Yes %			
Diabetes mellitus	40%	43%	37%
Hypertension	88%	84%	91%
Coronary artery disease	26%	28%	23%
Heart failure	19%	17%	21%
Primary renal disease, %			
Glomerular	11%	8.8%	13%
Tubulo-interstitial	10%	11%	9.3%
Diabetes	23%	24%	22%
Renovascular	35%	30%	41%
Other systemic disease	3.6%	3.6%	3.7%
Hereditary	4.4%	3.2%	5.7%
Miscellaneous	13%	19%	6.1%
Clinical measurements			
Current smoking, Yes %	7.0%	6.9%	7.0%
Systolic blood pressure mmHg, median (IQR)	146 (131, 160)	145 (130, 160)	148 (131, 160)
BMI kg/m^2, median (IQR)	27.7 (24.5, 31.2)	28.3 (24.7, 32.1)	27.0 (24.4, 30.2)
N of eGFR values per person, median (IQR)	5 (2, 8)	4 (2, 6)	7 (3, 9)
Laboratory measurements, median (IQR)			
eGFR CKD-EPI, ml/min/1.73m^{2*}	17.7 (14.6, 20.8)	17.8 (14.3, 21.1)	17.5 (15.1, 20.5)
eGFR slope, ml/min/1.73m^{2*} per year (95% CI)	–2.01 (–2.29, –1.73)	–1.97 (–1.73, –1.00)	–2.34 (–2.68, –1.99)
eGFR slope, % change per year (95% CI)	–14.6% (–16.7, –12.5)	–11.0% (–14.0, –7.9)	–16.5% (–19.4, –13.6)

ACR, mg/mmol	42 (8; 173)	44 (6; 174)	39 (10; 169)
Calcium, mmol/L	2.27 (2.17, 2.37)	2.27 (2.15, 2.37)	2.27 (2.19, 2.38)
Phosphate, mmol/L	1.29 (1.12, 1.48)	1.27 (1.13, 1.44)	1.30 (1.11, 1.50)
Parathyroid hormone, pmol/L	16 (10, 23)	16 (11, 25)	15 (9, 22)
Total cholesterol, mmol/l	4.60 (3.80, 5.50)	4.59 (3.90, 5.50)	4.60 (3.80, 5.50)
Kidney Failure Risk Equation (KFRE), median (IQR) *			
4-variable, 2-year	15.3% (5.8, 32.9)	15.3% (7.3, 31.7)	15.7% (4.6, 34.9)
8-variable, 2-year	18.5% (7.2, 34.3)	18.5% (8.1, 33.0)	18.3% (6.4, 37.1)
4-variable, 5-year	47.5% (20.6, 78.7)	47.4% (25.4, 77.1)	48.3% (16.6, 81.0)
8-variable, 5-year	60.1% (28.4, 84.9)	60.1% (31.6, 83.5)	59.8% (26.3, 87.6)

Table 8. 1 Baseline characteristics of participants included in this analysis

The annualised eGFR slopes were calculated without proteins and unadjusted for any independent variables using generalised linear mixed effects modelling. ** the KFRE predicts the risk of developing kidney failure within 2 and 5 years using the 4 predictors age, sex, eGFR and log(ACR) or 8 predictors age, sex, eGFR, log(ACR), serum calcium, serum phosphate, serum bicarbonate and serum albumin.(92) Abbreviations: **BMI**, body mass index; **eGFR**, estimated glomerular filtration rate; **ACR**, albumin: creatinine ratio.

8.1.2 Individual proteins primary analysis

Of 184 proteins, 11 proteins did not reach assay level of detection (liver carboxylesterase 1, neutrophil defensin 1, prolyl endopeptidase, integrin alpha-M, neutrophil gelatinase-associated lipocalin, latent-transforming growth factor beta-binding protein 2, platelet-activating factor acetyl hydrolase, superoxide dismutase 1, and uromodulin). As uromodulin and superoxide dismutase 1 are expected to decrease in CKD,(279–281) they may have a plausible association with kidney function decline and were not removed from the analysis. Therefore, the total number of proteins for analysis was 175.

The intra- and inter-assay Coefficient of Variance (CV) for the two protein panels (Target 96 cardiometabolic and Target 96 cardiovascular II) for the discovery sub-cohort was 6 – 7% and 12 – 19% respectively, and 5 – 6% (intra-assay) and 15% (inter-assay) for the validation sub-cohort. An intra-assay CV of < 15% and an inter-assay CV of < 25% are considered acceptable by Olink®. Laboratory validation of the protein panels can be found at https://olink.com/resources-support/document-download-center/.

In the discovery sub-cohort, 78 proteins affected eGFR slope positively, but none had evidence of an association using the threshold of P_{FDR}<0.05 (Figure 8. 2). Of the 97 proteins showing a negative relationship with eGFR slope, three were taken to validation: Receptor-type tyrosine-protein phosphatase S (PTPRS), Insulin-like growth factor binding protein 6 (IGFBP6) and Procollagen C-endopeptidase enhancer 1 (PCOLCE). PTPRS was found to have the largest effect on eGFR decline (-15.5% per year per protein concentration doubling; 95% CI -23.5, -7.5) which was sustained in the validation analysis (-11.4%; 95% CI -19.7, -4.5); see Figure 8. 3. IGFBP6 demonstrated smaller albeit more consistent effect sizes for the discovery

(-7.7%; 95% CI -12.1, -3.3) and validation sub-cohorts (-7.0%; 95% CI -12.0, -2.5) compared to PTPRS. Although PCOLCE was identified in the discovery analysis, it was found to have a smaller effect size in the validation analysis that bordered on null (discovery -9.4%; 95% CI -13.2, -5.5, versus validation -3.9%; 95% CI -8.5, -0.5) and so failed to be validated. Slope estimates for the discovery primary adjusted analysis are shown in Figure 8. 3 for all proteins.

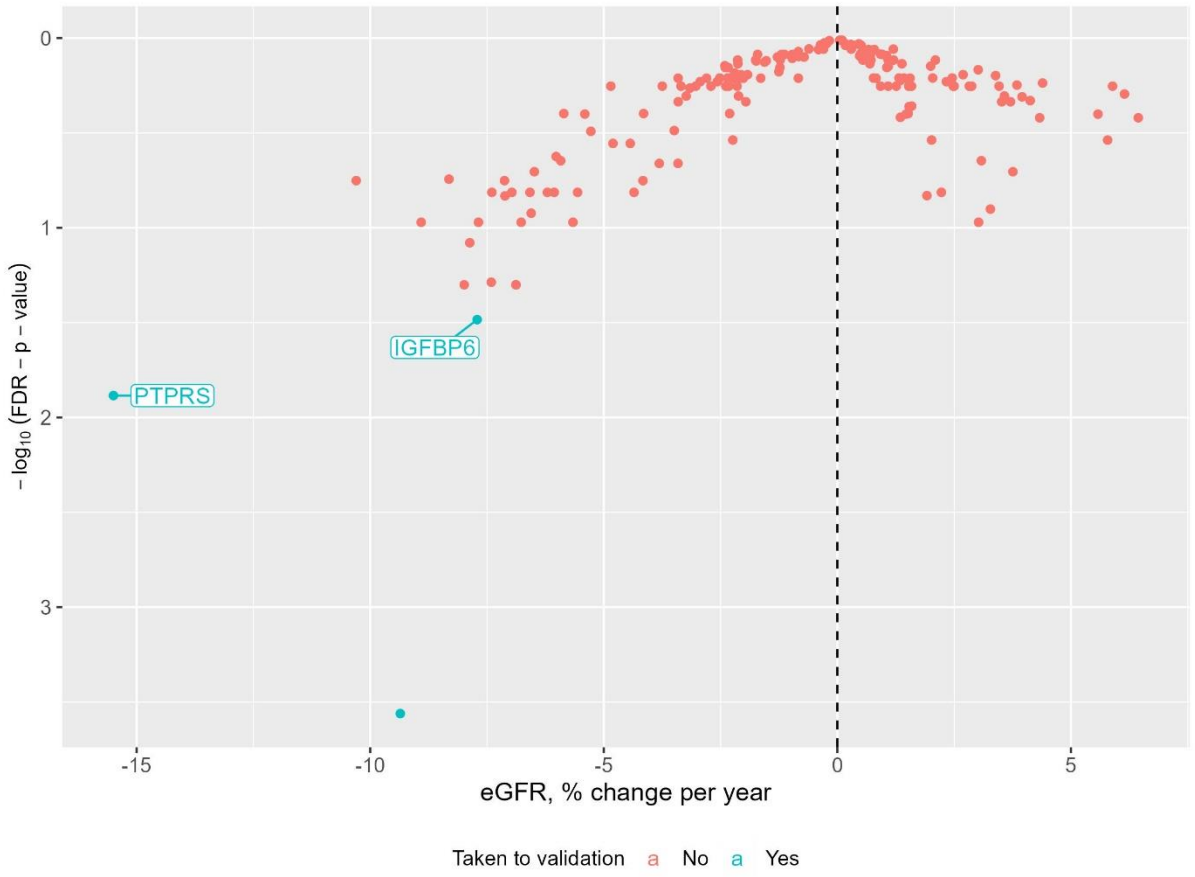


Figure 8. 2 Volcano plot of protein associations with eGFR slope for the discovery sub-cohort

Labelled proteins are those that replicated successfully in the primary analysis (adjusted for age, sex, country, diabetes mellitus status, systolic blood pressure, albumin: creatinine ratio, primary renal disease, renin-angiotensin aldosterone inhibitors and β-blockers). The unlabelled protein is PCOLCE which did not validate successfully. Coefficients are for the discovery sub-cohort. Abbreviations: **eGFR**, estimated glomerular filtration rate; **PTPRS**, Receptor type tyrosine-protein phosphatase S; **IGFBP6**, Insulin-like growth factor binding protein.

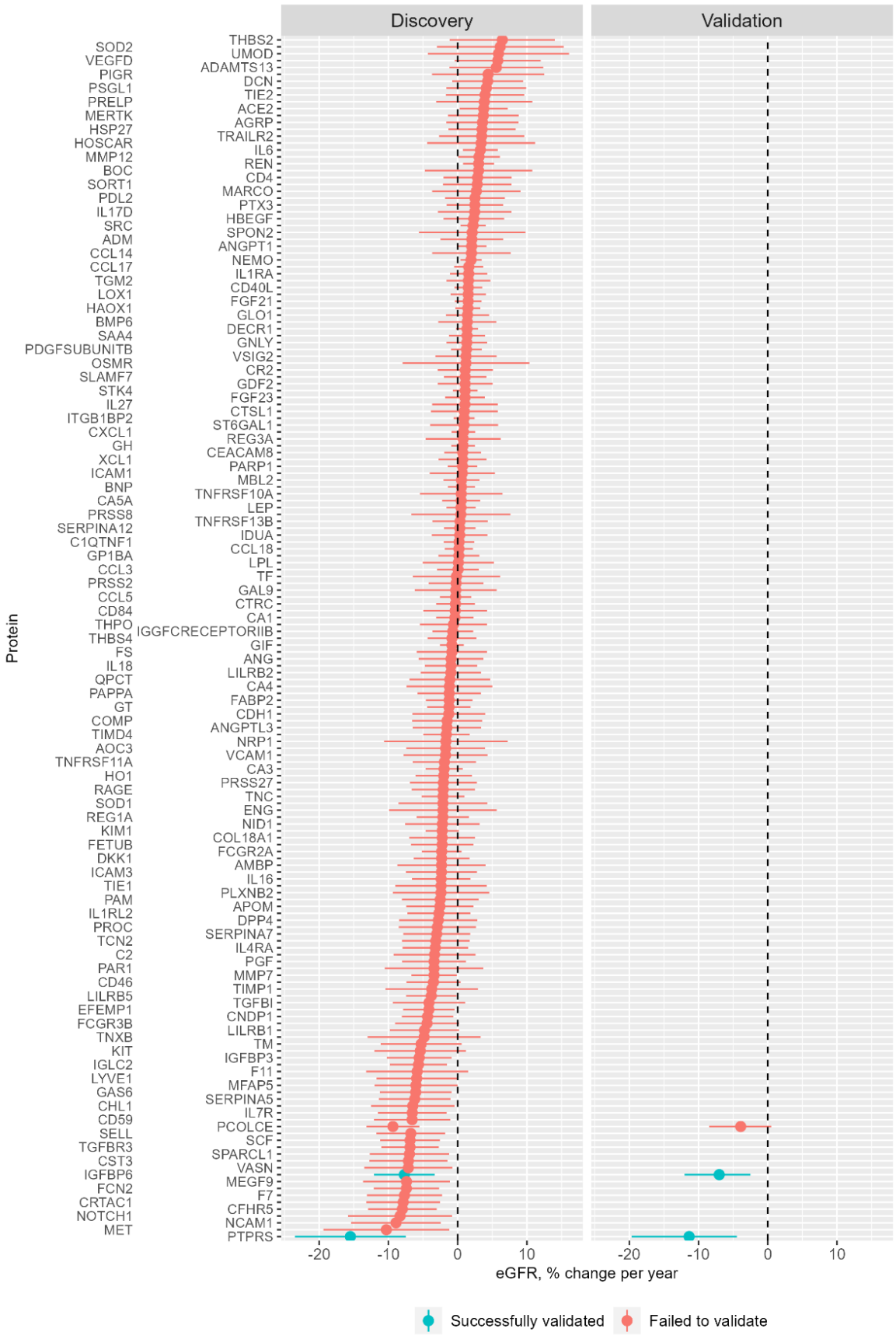


Figure 8. 3 Forest plot of protein associations with eGFR slope

Solid circles represent the slope estimates for each protein and the horizontal bars represent the 95% confidence intervals. Note that the effect of GFR change is per protein concentration doubling as 1 NPX change in protein is relative and uninterpretable. Estimates are shown for the primary analysis model 4 (adjusted for age, sex, country, diabetes mellitus status, systolic blood pressure, albumin: creatinine ratio, primary renal disease, renin-angiotensin aldosterone inhibitors and β-blockers). Proteins that failed to discover were not tested in the validation sub-cohort and their slope estimates are, therefore, unavailable. Abbreviations: eGFR, estimated glomerular filtration rate; **PCOLCE**, Procollagen C-endopeptidase enhancer 1; **IGFBP6**, Insulin-like growth factor-binding protein 6; **PTPRS**, Receptor-type tyrosine-protein phosphatase S. Other protein abbreviations are available in the Appendix 2.

For model 3, the Ficolin-2-slope (FCN2) estimates were -7.6% (95% CI -12.1, -3.0) in discovery and -4.4% (95% CI -8.4, -0.4) in validation. FCN-2 was associated with eGFR decline but not after additional adjustment for medications i.e., the primary analysis model 4. Slope effect estimates did not differ substantially using different model adjustments (Figure 8. 4).

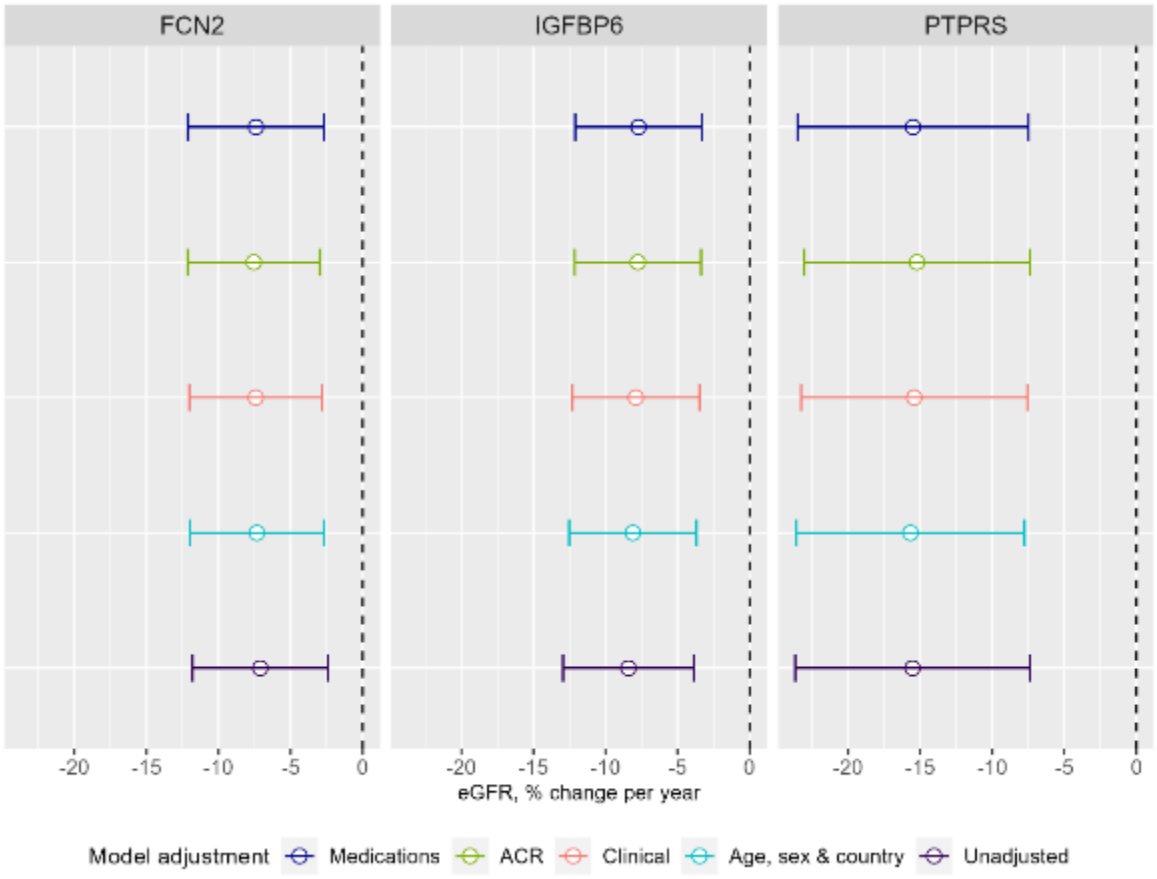


Figure 8. 4 GLM model adjustment comparison

Hollow circles represent the slope estimates and the horizontal bars represent the 95% confidence intervals (CI). Note that the effect of GFR change is per protein concentration doubling as 1 NPX change in protein is relative and uninterpretable. CIs do not cross the null (0%) which would suggest an association but the CIs could not be adjusted for Benjamini-Hochberg multiple testing correction unlike the *P*-values hence the discrepancy that the CI for FCN-2 does cross the null for the Medications-adjusted model (model 4) but was still considered as having no evidence of an association with eGFR slope for this model adjustment. Minimal adjustment (model 1): age, sex and country. Clinical adjustment (model 2): model 1 + primary renal disease, diabetes mellitus status and systolic blood pressure. ACR adjustment (model 3): model 2 + ACR. Medications (model 4): model 3 + renin-angiotensin aldosterone system inhibitors and β-blockers. Abbreviations: **GLMM**, generalised linear mixed effects model; **eGFR**, estimated glomerular filtration rate; **ACR**, albumin: creatinine ratio; **IGFBP6**, Insulin-like growth factor binding protein 6; **PTPRS**, Receptor-type tyrosine-protein phosphatase S; **FCN2**, Ficolin-2.

Inspection of the predicted slope plots in Figure 8. 5 suggests that lower protein levels were universally associated with less steep eGFR decline i.e., lower levels of protein were protective. For IGFBP6 and PTPRS, the baseline eGFR was lower at higher protein abundance in addition to being associated with a steeper slope (Figure 8. 5).

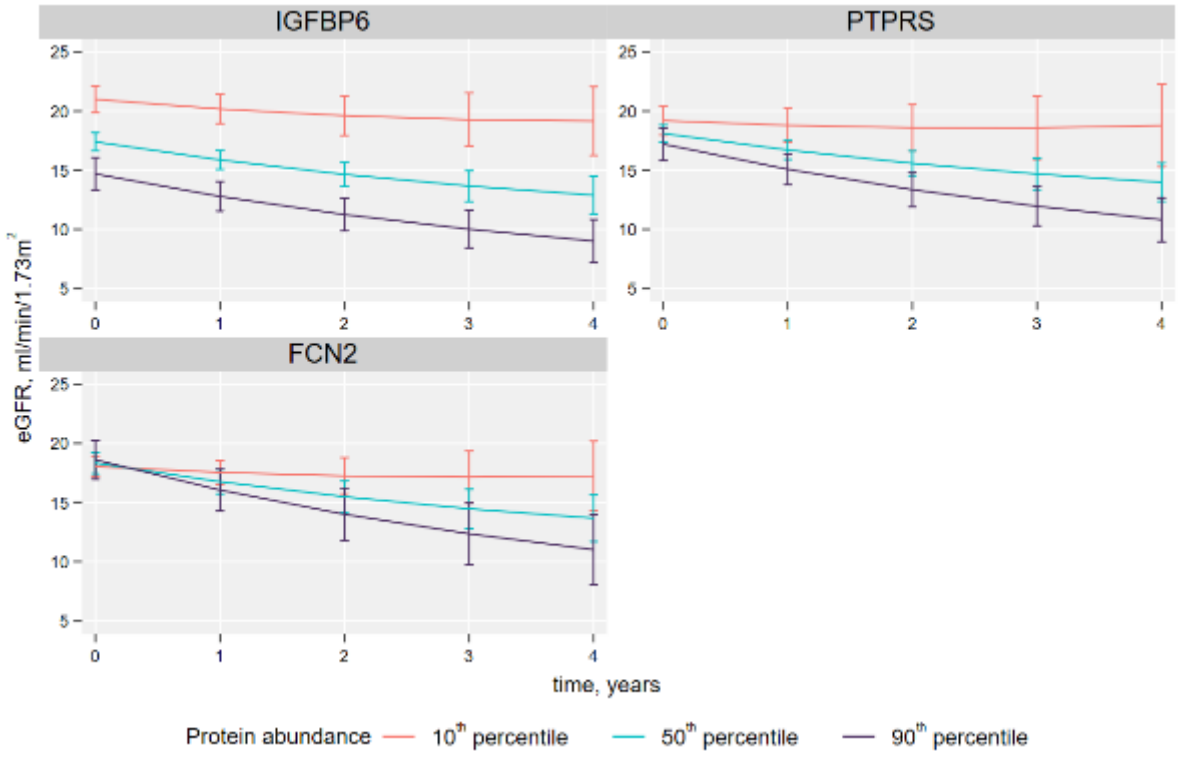


Figure 8. 5 eGFR slope for each successfully validated protein in the primary analysis

Horizontal lines represent the slope and vertical bars show the 95% confidence limits. Note that the effect of GFR change is per protein concentration doubling as 1 NPX change in protein is relative and uninterpretable. Estimates are for model 4 (discovery sub-cohort) although only IGFBP6 and PTPRS were successfully validated in that adjusted analysis (FCN2 was only validated in model 3). Time was truncated to 4 years because of a questionable upstroke in the predicted eGFR slope after this time, likely due to sparse data > 4 years (see Figure 8. 1). Abbreviations: **eGFR**, estimated glomerular filtration rate; **IGFBP6**, Insulin-like growth factor binding protein 6; **PTPRS**, Receptor-type tyrosine-protein phosphatase S; **FCN2**, Ficolin-2.

Positive correlations between individual proteins were observed (Figure 8. 6) and a congruous positive correlation was seen between baseline and repeatedly measured eGFR. The correlation of proteins with ACR was weakly-positive. Correlations between proteins and other traditional cardiovascular risk factors, age, sex, systolic blood pressure, diabetes mellitus and important protective cardiovascular drugs were weak perhaps explaining similar magnitude protein-eGFR associations estimated across model adjustments with and without these risk factors (Figure 8. 4).

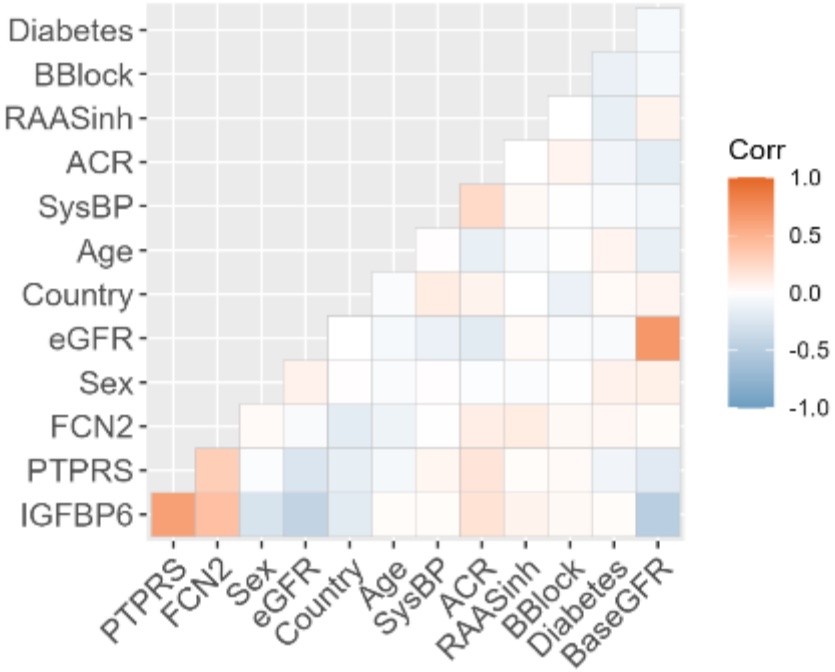


Figure 8. 6 Correlation matrix of individual proteins and clinical covariables

The heatmap shows the correlation between proteins and clinical variables. 'BaseGFR' is the eGFR at the first study visit whereas 'eGFR' refers to the longitudinal eGFR over time. Abbreviations: **Corr,** correlation; **BaseGFR**, baseline eGFR; **BBlock**, β-blockers; **RAASinh**, renin-angiotensin aldosterone system inhibitors; **ACR**, (urinary) albumin: creatinine ratio; **SysBP**, systolic blood pressure; **eGFR**, estimated glomerular filtration rate; **IGFBP6**, Insulin-like growth factor binding protein 6; **PTPRS**, Receptor-type tyrosine-protein phosphatase S; **FCN2**, Ficolin-2.

8.1.3 Follow-up time and censoring events

Follow-up time averaged one year longer in the validation sub-cohort (discovery 2.6 years; IQR 1.2, 4.3 compared to validation 3.6 years; IQR 1.9, 4.0). The reason for end of follow-up was KRT initiation in 40% and 34% of participants, for discovery and validation respectively. In the discovery sub-cohort 22% of participants died compared to 26% in the validation set, and more participants were lost to follow-up for discovery (16% versus 6%). Figure 8. 7 shows that eGFR steadily declined for those who went on to initiate KRT but was more variable over time for other reasons of loss to follow-up. This was corroborated in a joint model accounting for KRT initiation (Figure 8. 8). In addition to a steeper decline, baseline eGFR was also lower for those that initiated KRT.

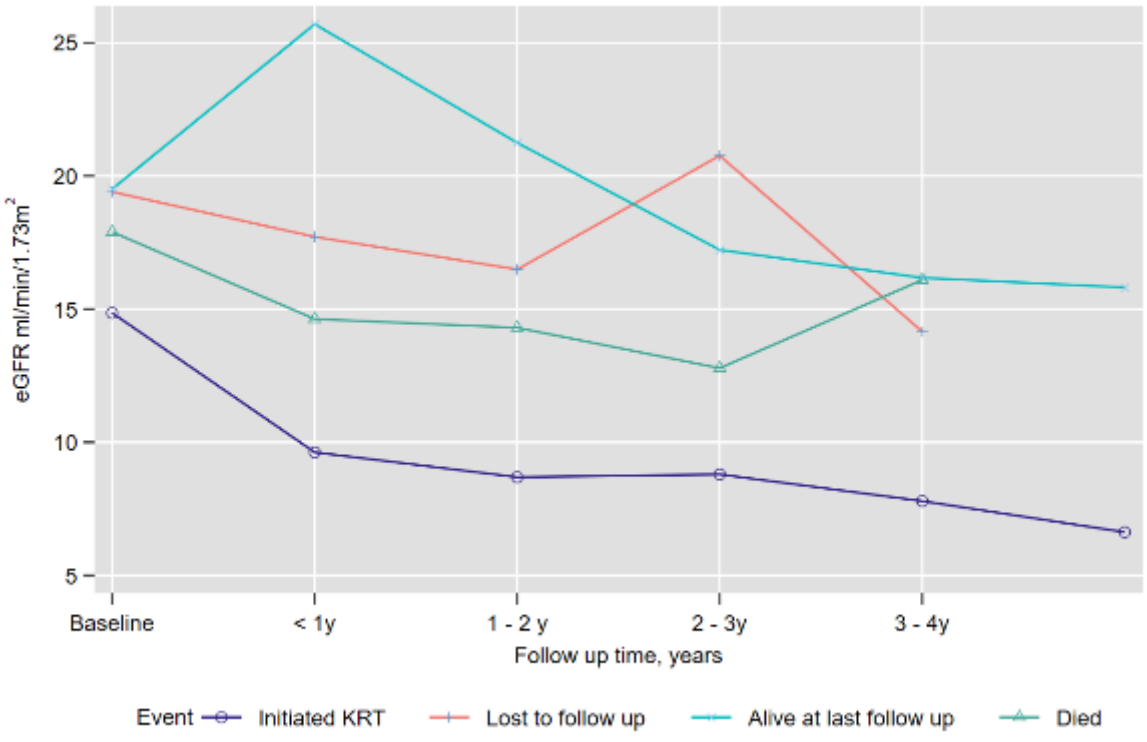


Figure 8. 7 Median eGFR over time for different censoring events

Abbreviations: **eGFR**, estimated glomerular filtration rate; **KRT**, kidney replacement therapy; **y**, years.

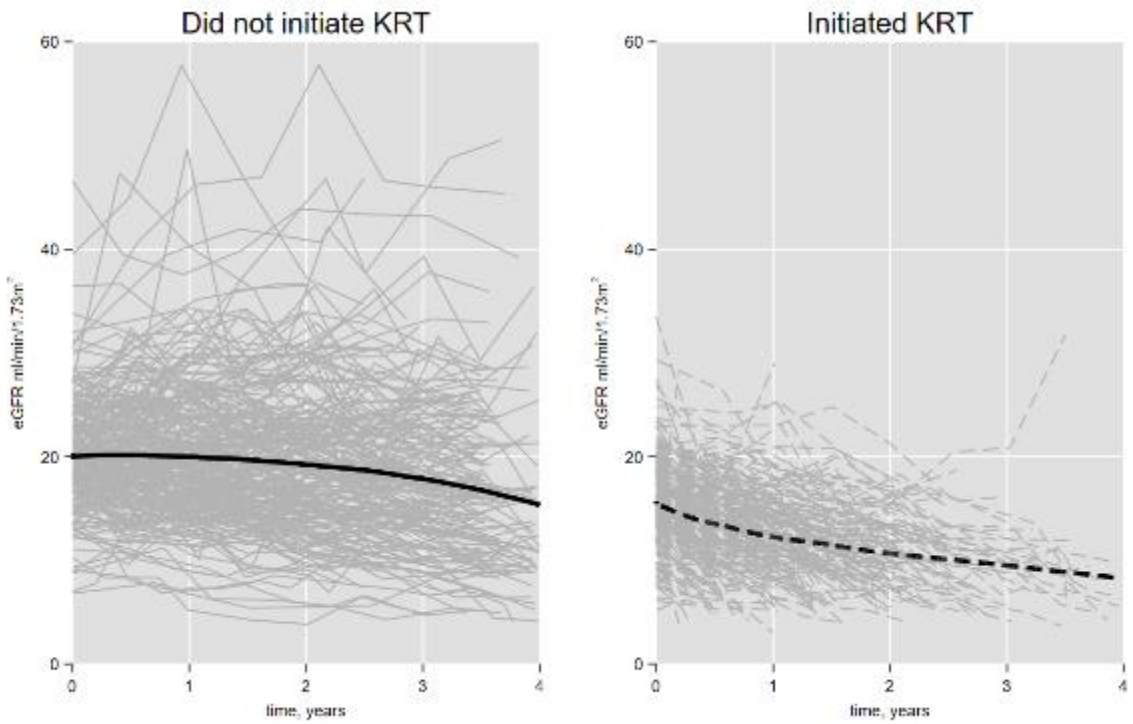


Figure 8. 8 eGFR slope accounting for KRT initiation in a joint model

Those who did not initiate KRT only included people who did not die as those that did would be expected to be different to people who did not die. The Stata `stjmgraph` command was used to graph this joint model which does not allow for competing risks of other potential outcomes. No protein was included. The grey lines show individual slopes for each person whilst the black lines show the average population slope. Abbreviations: **eGFR**, estimated glomerular filtration rate; **FU**, follow up; **KRT**, kidney replacement therapy.

8.1.4 Joint model sensitivity analysis

Slope estimates derived using the joint model were very similar compared to the primary GLMM analysis suggesting a negligible bias introduced by informative censoring and competing risks (see Figure 8. 9).

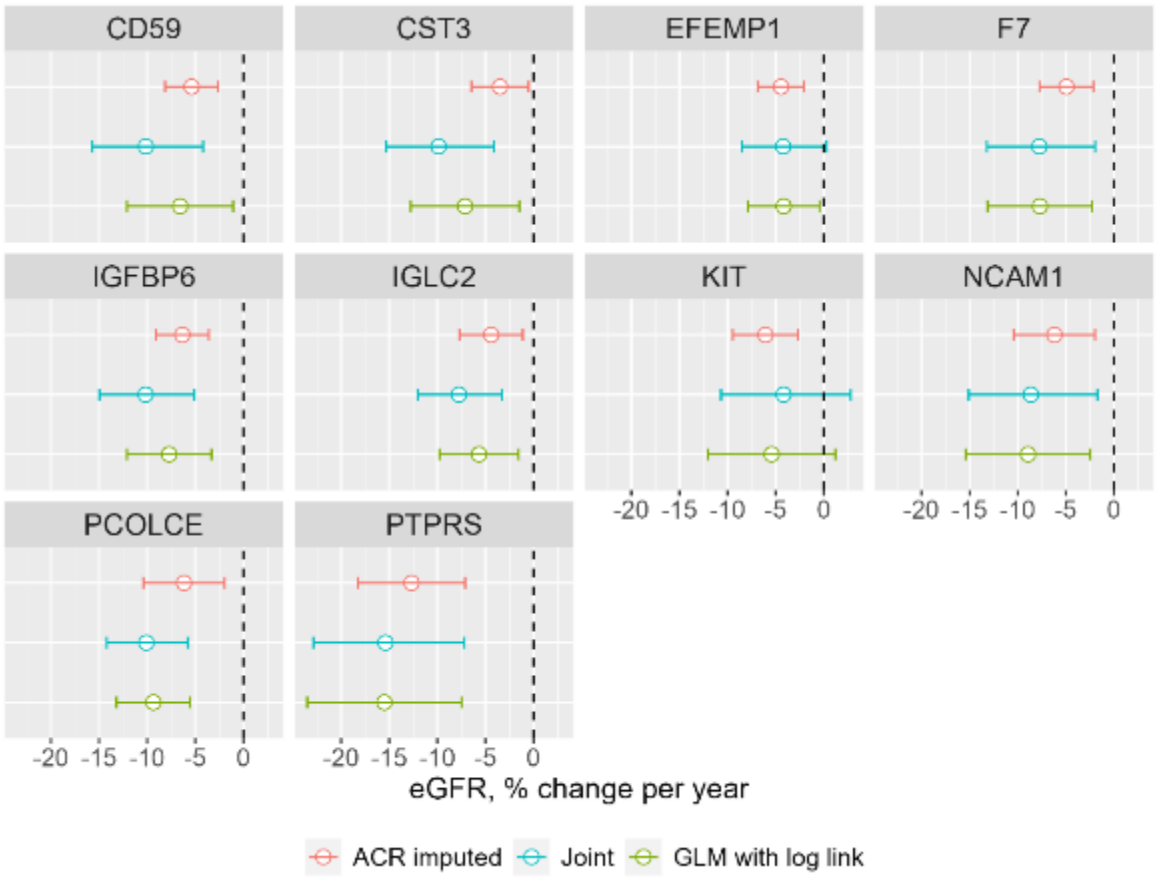


Figure 8. 9 Comparison of slope estimates for GLM, imputation and joint models

Hollow circles represent the slope estimates, and the horizontal bars represent the 95% confidence intervals. Note that the effect of GFR change is per protein concentration doubling as 1 NPX change in protein is relative and uninterpretable. CIs were not corrected for multiple testing so may not cross the null (0%). More proteins were validated in the joint and ACR imputed models compared to the primary GLM model. Abbreviations: **GLM**, generalised linear mixed effects; **EFEMP1**, EGF-containing fibulin-like extracellular matrix protein 1; **IGFBP6**, Insulin-like growth factor binding protein 6; **CD59**, Cluster of differentiation 59; **CST3**, Cystatin 3; **IGLC2**, Immunoglobulin lambda-2 chain C regions; **NCAM1**, Neural cell adhesion molecule 1; **PCOLCE**, Procollagen C-endopeptidase enhancer 1; **PTPRS**, Receptor-type tyrosine-protein phosphatase S; **F7**, Factor 7; **KIT**, Mast/ stell cell receptor Kit.

8.1.5 Multiple imputation sensitivity analysis

Patients with complete and in-complete data are described in Table 8. 2. Baseline characteristics were very similar except that differences in country of residence are notable. Even after adjustment for primary renal disease (stricter monitoring of ACR may have been expected in individuals with glomerulonephritides or diabetic nephropathy diagnoses) or country (as a proxy of local testing practices), and other variables, the outcome (slope) did not increase the odds of ACR being unobserved (OR 0.71; 95% CI 0.42, 1.18; *P* = 0.187).

The sensitivity analysis using multiple imputation of ACR found additional proteins to be associated with eGFR loss and all proteins identified in the model 4 CCA were also associated with eGFR slope in the MI analysis. These were: EGF-containing fibulin-like extracellular matrix protein 1 (EFEMP1), Cluster of differentiation 59 (CD59), Cystatin 3 (CST3), Immunoglobulin lambda-2 chain C regions (IGLC2), Neural cell adhesion molecule 1 (NCAM1), Procollagen C-endopeptidase enhancer 1 (PCOLCE), Receptor-type tyrosine-protein phosphatase S (PTPRS), Factor 7 (F7), Mast/ stell cell receptor Kit (KIT). The eGFR slopes were most often less steep (eight proteins) but consistent (overlapping confidence intervals) with the primary GLMM analysis (shown in Figure 8. 9).

	Complete cases, n = 501	**Incomplete cases**, n = 432
Age in years, median (IQR)	75 (70, 80)	77 (72, 82)
Female sex, %	33%	39%
Country, %		
Germany	8.5%	20%
United Kingdom	29%	67%
Poland	12%	1.6%
Sweden	51%	11%
Comorbidities, Yes %		
Diabetes mellitus	40%	43%
Hypertension	88%	83%
Coronary artery disease	36%	23%
Heart failure	19%	14%
Primary renal disease, %		
Glomerular	11%	9.2%
Tubulo-interstitial	9.9%	9.2%
Diabetes	23%	20%
Renovascular	35%	31%
Other systemic disease	3.7%	1.9%
Hereditary	4.3%	3.8%
Miscellaneous	13%	25%
Clinical measurements		
Current smoking, Yes %	7.1%	8.3%
Systolic blood pressure mmHg, median (IQR)	146 (131, 160)	144 (130, 160)
BMI kg/m^2, median (IQR)	27.6 (24.5, 31.2)	28.7 (25.4, 32.9)
N of eGFR values per person, median (IQR)	7 (5, 9)	5 (4, 8)
Total N of eGFR values	2,494	1,978
Laboratory measurements, median (IQR)		
eGFR CKD-EPI, ml/min/1.73m^2	17.6 (14.7, 20.8)	17.4 (13.7, 21)
Calcium, mmol/L	2.27 (2.17, 2.37)	2.29 (2.18, 2.38)
Phosphate, mmol/L	1.29 (1.12, 1.46)	1.27 (1.10, 1.45)
Parathyroid hormone, pmol/L	16 (10, 23)	16 (8, 25)
Total cholesterol, mmol/L	4.60 (3.80, 5.50)	4.27 (3.40, 5.27)

Table 8. 2 Comparison of baseline characteristics of participants with and without complete data at baseline

Abbreviations: **BMI**, body mass index; **eGFR**, estimated glomerular filtration rate; **ACR**, albumin: creatinine ratio; **IQR**, interquartile range.

8.1.6 Biological pathway analysis

Analysis of the 35 individual differentially expressed proteins meeting a liberal threshold of $P < 0.05$ identified 373 biological pathways in the Reactome knowledgebase. Only about 50% of the human genome is represented in the database to date.(232) As such, some proteins could not be found despite using alternative annotations (Vasorin, Transforming growth factor β receptor type 3, Cartilage acidic protein 1 and Multiple epidermal growth factor-like domains protein 9).[27] There was evidence against nine pathways were not over-represented ($P_{FDR} < 0.05$) (Table 8. 3), many of which are ubiquitous intracellular processes.

[27] Personal communication with Reactome – Reactome's 'interactor' catalogue does contain these proteins, but their over-representation database does not.

Pathway name	Proteins found	Total*	Ratio#	FDR *P*-value
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	6	124	0.010	0.001
Post-translational protein phosphorylation	5	107	0.009	0.005
Extracellular matrix organization	7	300	0.026	0.006
Transport of gamma-carboxylated protein precursors from the endoplasmic reticulum to the Golgi apparatus	2	9	0.001	0.031
Removal of amino-terminal pro-peptides from gamma-carboxylated proteins	2	10	0.001	0.031
Gamma-carboxylation of protein precursors	2	10	0.001	0.031
PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling	4	118	0.010	0.031
Gamma-carboxylation, transport, and amino-terminal cleavage of proteins	2	11	0.001	0.031
Negative regulation of the PI3K/AKT network	4	125	0.012	0.031

Table 8. 3 Biological pathways identified using over-representation analysis

A search for the analysed proteins was undertaken using the Reactome knowledgebase (www.reactome.org, Creative Commons BY 4.0 license). * proteins that are known to be involved in the pathway and are catalogued within reactome. # proteins found divided by all proteins in the pathway. PTPRS protein is involved in the 'extracellular matrix organisation' pathway and IGFBP6 is involved in 'Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)'. Abbreviations: **FDR**, false discovery rate; **PI3K**, phosphoinositide 3-kinase; **PI5P**, phosphatidylinositol 5-phosphate; **PP2A**, protein phosphatase 2; **IER3**, immediate early response 3; **AKT**, protein kinase B.

The proteins PTPRS and IGFBP6 contributed the largest variance to each first principal component for the *Extracellular Matrix Organisation* and *Regulation of IGF Transport and Uptake by IGFBPs* pathways, respectively (Figure 8. 10 and Figure 8. 11). Visualisation of scree plots Figure 8. 12 and Figure 8. 13 demonstrated that the first three principal components represent most of the total variance. Predicted slope estimates are shown for the two pathways in Table 8. 4. The nett % change is notably lower than for the individual proteins PTPRS and IGFBP6. Figure 8. 14, a protein-protein interaction network, would suggest little evidence for an interaction between the proteins despite them belonging to the same pathways. Other, not analysed proteins in this analysis, may have stronger interactions with these proteins which might explain the small eGFR slopes associated with these principal components which only include the analysed proteins in the pathway. Furthermore, the first principal component eGFR slope association was negative whereas eGFR slope was positive for the second and third principal components. This is due to the relative contribution of different proteins to these principal components which may have opposing effects on eGFR decline.

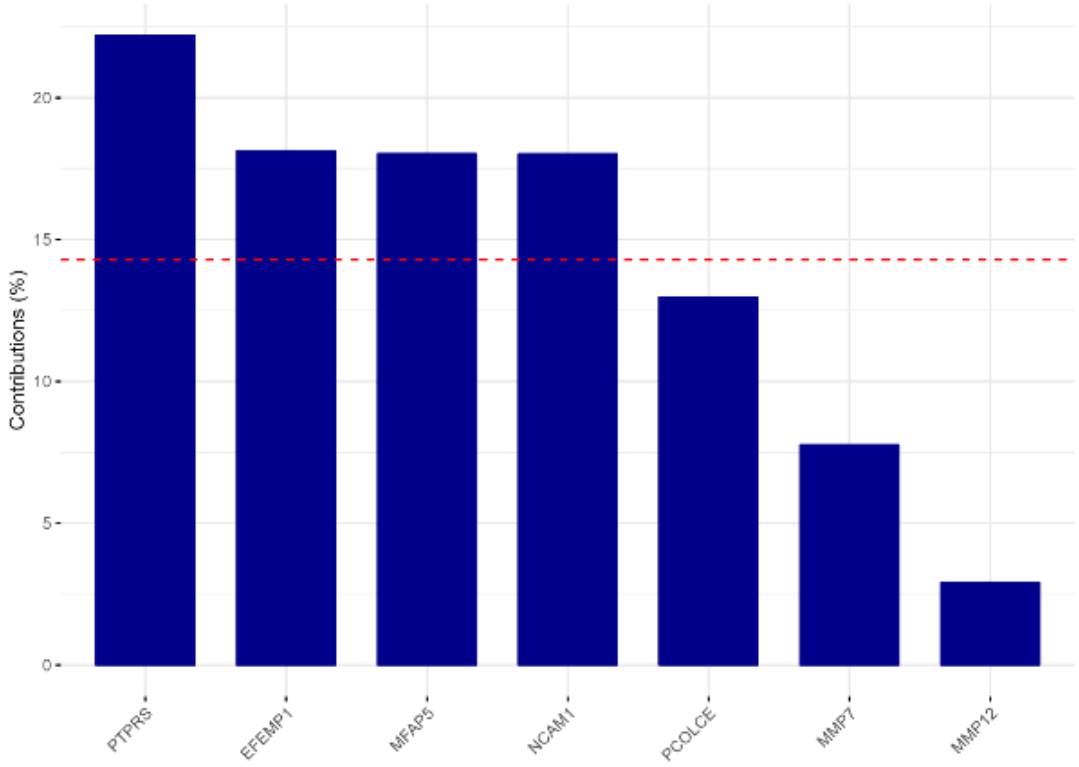


Figure 8. 10 Contributions of individual proteins on principal component 1 (extracellular matrix organisation)

The dashed red line is the mean contribution. Abbreviations: **PTPRS**, Receptor-type tyrosine-protein phosphatase S; **EFEMP1**, EGF-containing fibulin-like extracellular matrix protein 1; **MFAP5**, Microfibrillar-associated protein 5; **NCAM1**, Neural cell adhesion molecule 1; **PCOLCE**, Procollagen C-endopeptidase enhancer 1; **MMP7**, Matrix metalloproteinase-7; **MMP12**, Matrix metalloproteinase-12.

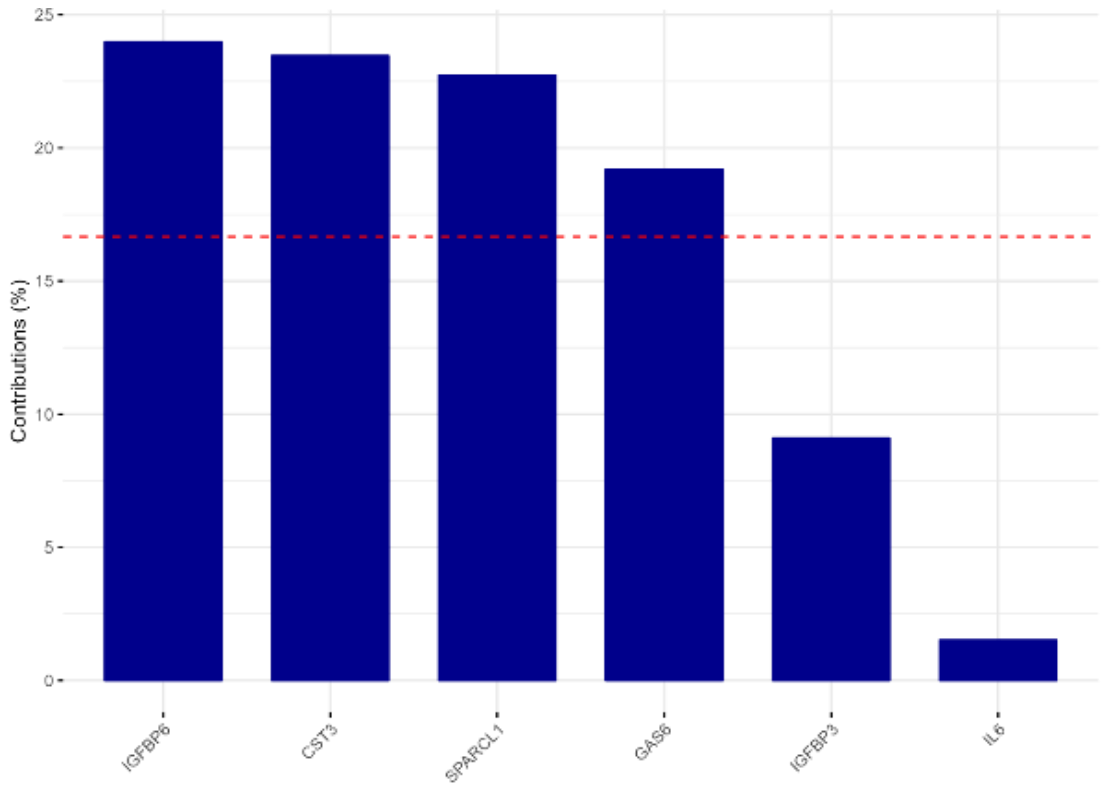


Figure 8. 11 Contributions of individual proteins on the first principal component (Regulation of IGF transport and uptake by binding proteins)

The red dashed line represents the average contribution. Abbreviations: **IGFBP6**, Insulin-like growth factor-binding protein 6; **CST3**, Cystatin-C; **SPARCl1**, SPARC-like protein 1; **GAS6**, Growth arrest-specific protein 6; **IGFBP3**, Insulin-like growth factor-binding protein 3; **IL6**, Interleukin 6.

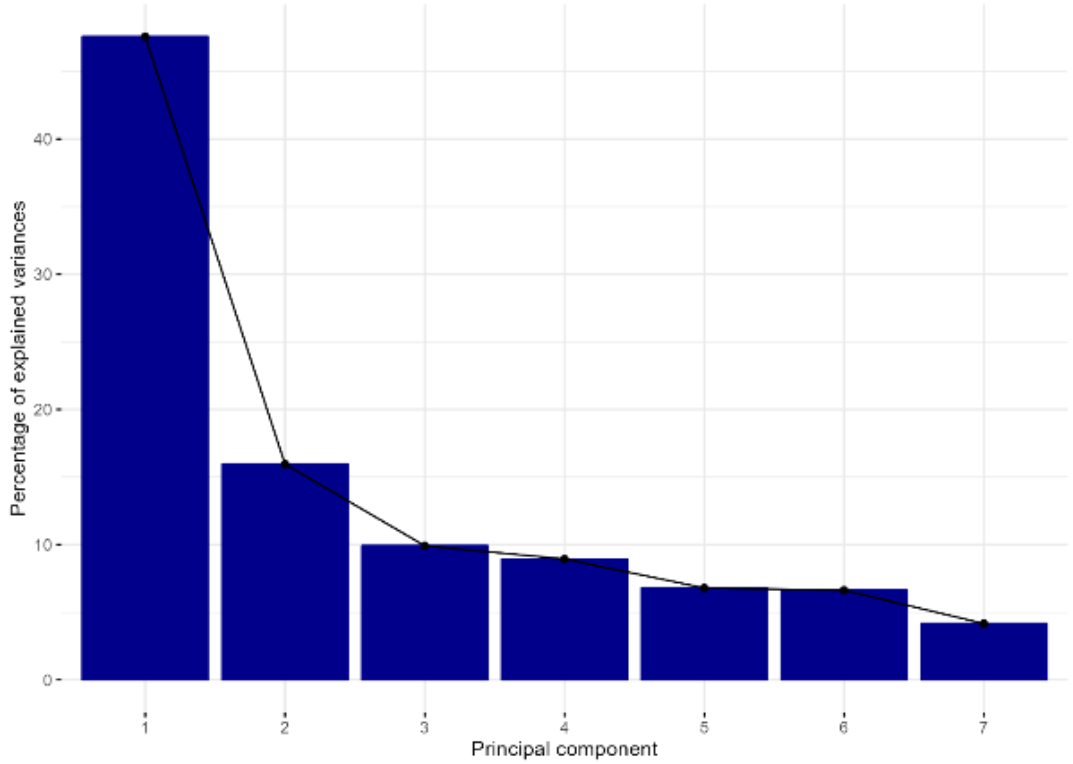


Figure 8. 12 Scree plot of principal components (extracellular matrix organisation)

The scree plot demonstrates that the variance explained by each principal component does not remarkably improve after the third principal component.

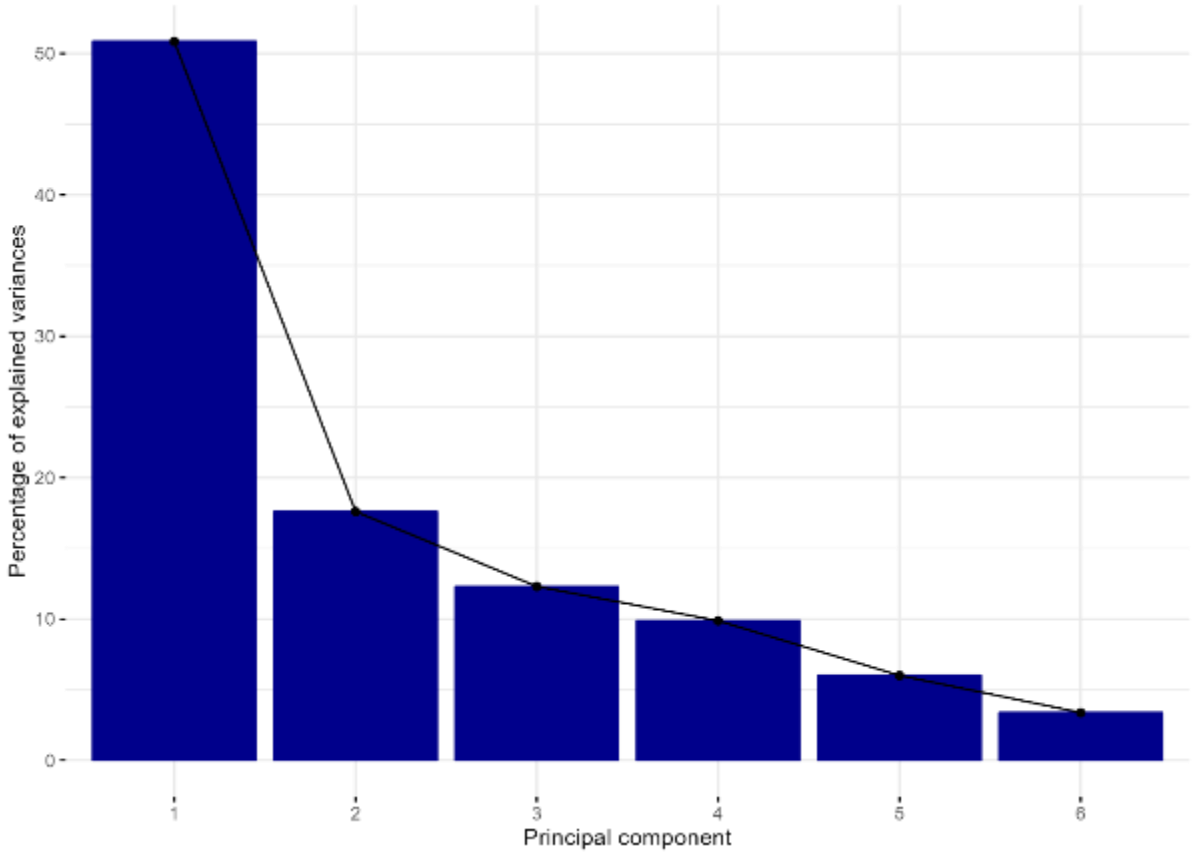


Figure 8. 13 Scree plot of principal components (Regulation of IGF transport and uptake by binding proteins)

The scree plot demonstrates that the variance explained by each principal component does not remarkably improve after the third principal component.

Pathway ***Principal component***	**eGFR % Change per year (95% CI)**	
	Discovery cohort	**Validation cohort**
Extracellular matrix organisation		
1	-2.3% (-3.8, -0.9)	-2.6% (-4.1, -1.1)
2	3.4% (0.5, 6.3)	1.4% (-1.5, 4.3)
3	2.2% (-1.3, 5.8)	-0.5% (-3.9, 3.0)
Regulation of IGF transport and uptake by binding proteins		
1	-2.2% (-3.6, -0.7)	-1.1% (-2.8, 0.5)
2	4.2% (1.2, 7.1)	0.5% (-1.8, 2.7)
3	2.0% (-1.3, 5.2)	0.7% (-2.3, 3.7)

Table 8. 4 Slope estimates for the first three principal components

For each biological protein, slope estimates are shown for the first three principal components. Abbreviations: **eGFR**, estimated glomerular filtration rate; **IGF**, Insulin-like growth factor; **CI**, confidence interval.

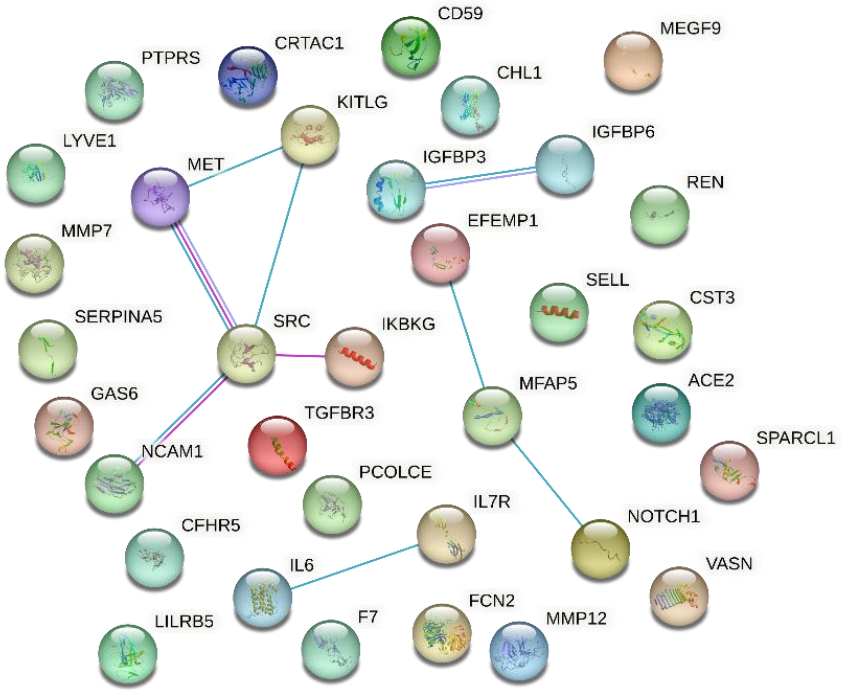


Figure 8. 14 Protein-protein interaction network

Protein-protein interaction network created using string-db.org (Creative Commons BY 4.0 licence) of the 35 proteins that were differentially expressed. Blue edges denote evidence of an interaction using curated databases and pink edges denote experimentally determined relationships. Node colours have no significance. Abbreviations: **ACE2**, Angiotensin-converting enzyme 2; **CD59**, CD59 glycoprotein; **CFHR5**, Complement factor H-related protein 5; **NCAM1**, Neural cell adhesion molecule L1-like protein; **CRTAC1**, Cartilage acidic protein 1; **CST3**, Cystatin-C; **F7**, Coagulation factor VII; **FCN2**, Ficolin-2; **GAS6**, Growth arrest-specific protein 6; **IGFBP3**, Insulin-like growth factor-binding protein 3; **IGFBP6**, Insulin-like growth factor-binding protein 6; **IL6**, Interleukin-6; **IL7R**, Interleukin-7 receptor subunit alpha; **KIT**, Mast/stem cell growth factor receptor Kit; **LILRB5**, Leukocyte immunoglobulin-like receptor subfamily B member 5; **LYVE1**, Lymphatic vessel endothelial hyaluronic acid receptor 1; **MEGF9**, Multiple epidermal growth factor-like domains protein 9; **MET**, Hepatocyte growth factor receptor; **MFAP5**, Microfibrillar-associated protein 5; **MMP12**, Matrix metalloproteinase-12; **MMP7**, Matrix metalloproteinase-7; **NCAM1**, Neural cell adhesion molecule 1; **IKBKG** (NEMO), NF-kappa-B essential modulator; **NOTCH1**, Neurogenic locus notch homolog protein 1; **PCOLCE**, Procollagen C-endopeptidase enhancer 1; **PTPRS**, Receptor-type tyrosine-protein phosphatase S; **REN**, Renin; **SELL**, L-selectin; **SREPINA5**, Plasma serine protease inhibitor; **SPARCL1**, Secreted protein acidic and rich in cysteine-like protein 1 (osteonectin); **SRC**, Proto-oncogene tyrosine-protein kinase Src; **VASN**, Vasorin.

8.1.7 Model checking and diagnostics

The validity of model assumptions, fully adjusted but before the inclusion of proteins, was checked by assessing patterns in plots of model parameters. A Quantile-Quantile plot, shown in Figure 8. 15, was used to check that the GLM model residuals were approximately normally distributed. Furthermore, heteroscedasticity[28] was excluded by plotting the residuals versus the fitted model values, shown in Figure 8. 16. For the multiple imputation sensitivity analysis, the imputed values were comparable with the observed values as shown in Figure 8. 17 and the imputed values varied randomly over successive iterations as visualised by plotting the mean of ACR for each imputed dataset, shown in Figure 8. 18. Together, this supports that MI produced valid potential ACR values.

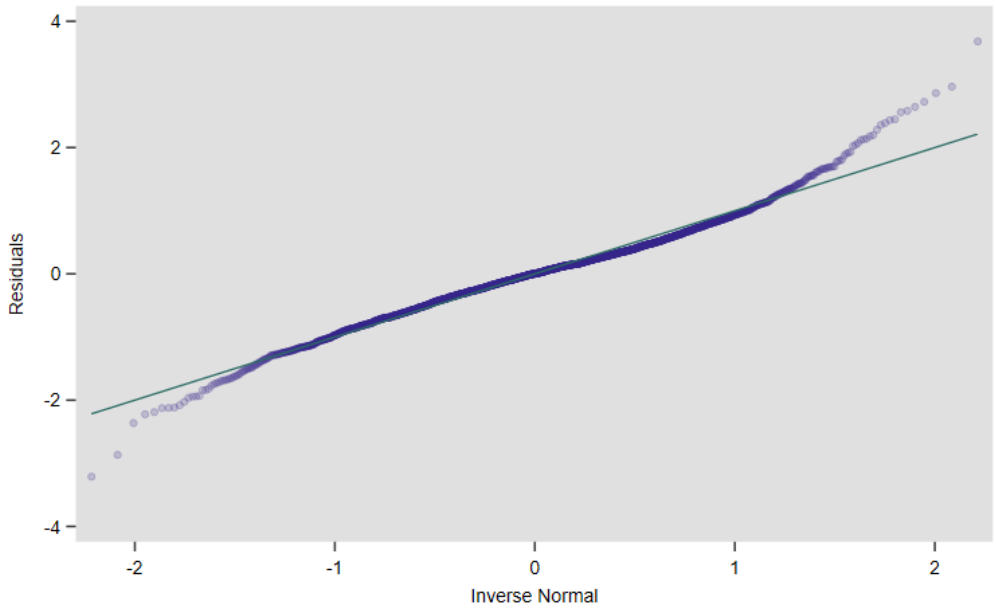


Figure 8. 15 Quantile-Quantile plot showing the distribution of model residuals

The residuals were approximately normally distributed as they generally follow the line of equality, except at the extremes where some curve was noted.

[28] Heteroscedasticity is an undesirable model attribute that biases model-based standard errors in the presence of inconstant variance between observations.

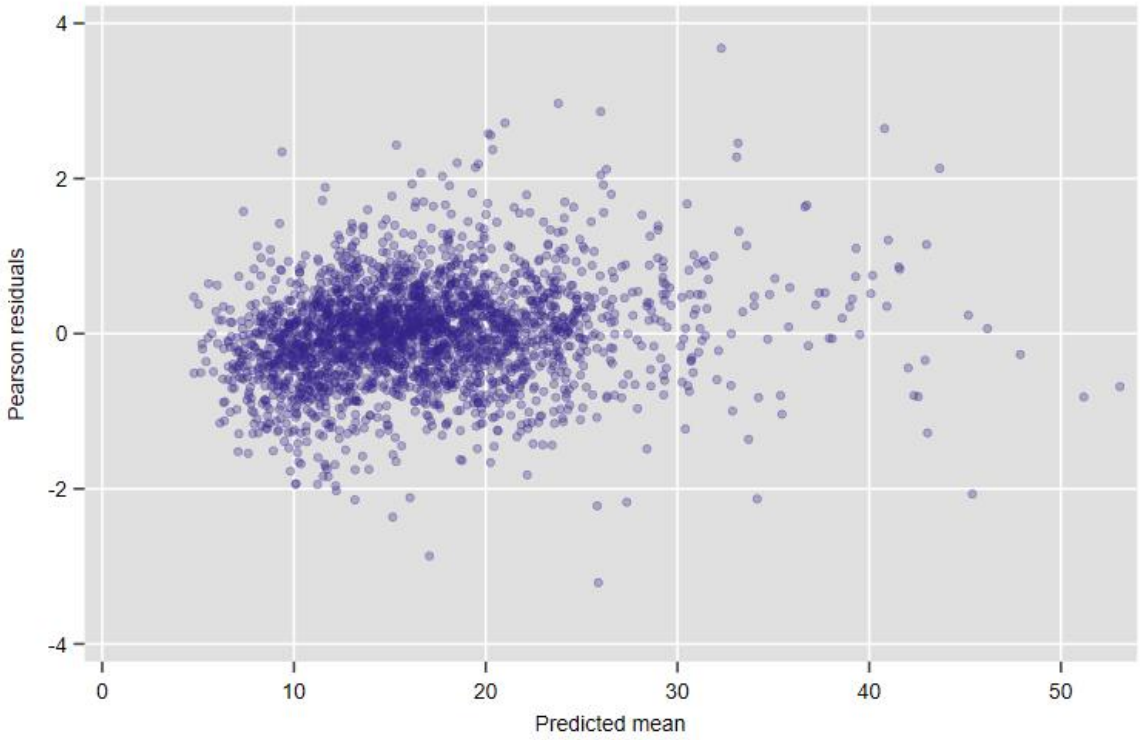


Figure 8. 16 Model residuals versus fitted values check for heteroscedasticity

Towards the left of the plot, points are grouped together in no particular pattern. Data points on the right are less dense but still do not display an obvious relationship between the model residuals and fitted values.

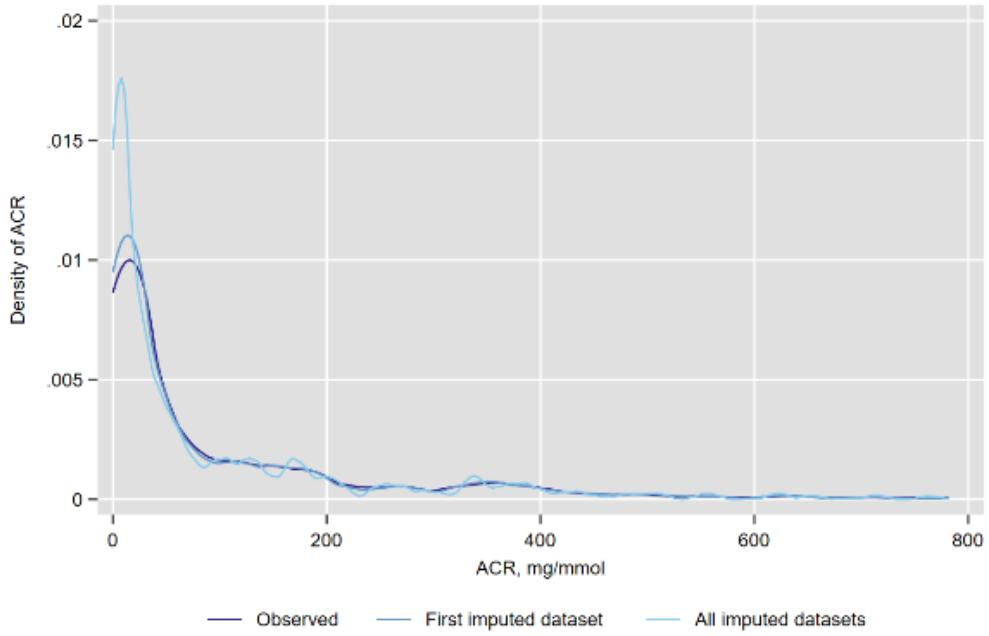


Figure 8. 17 Compatibility of imputed datasets with observed values of ACR

Observed ACR values are compared to the first imputed dataset and follow the same shape for all remaining imputed datasets. Abbreviations: **ACR**, albumin: creatinine ratio.

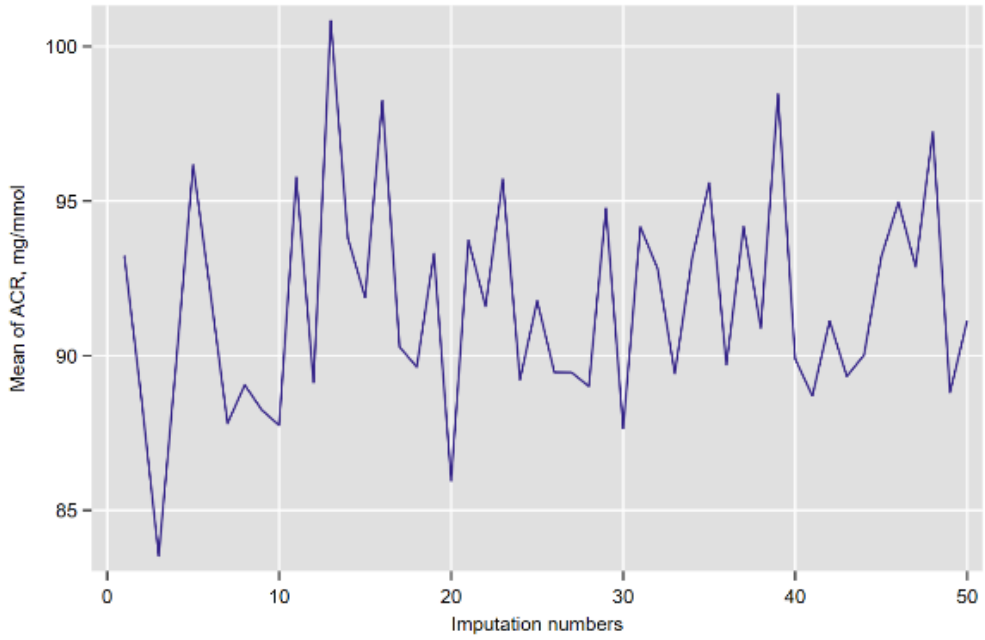


Figure 8. 18 Mean ACR for each imputed dataset

There was a wide and random variability in the mean ACR values for each imputed dataset confirming a correctly specified MI that was not biased towards a particularly low or high ACR value. Abbreviations: **ACR**, albumin: creatinine ratio.

8.2 DISCUSSION

Using data from the EQUAL study, two proteins – Receptor protein-tyrosine phosphatase sigma and Insulin-like growth factor binding protein 6 – were identified as having a sustained association with eGFR decline in two independent sub-cohorts of older Europeans with advanced CKD in the primary adjusted analysis including medications (model 4). After including adjustment for medications, Ficolin-2 no longer displayed an association unlike PTPRS and IGFBP6. Slope estimates were almost identical across all model adjustments perhaps indicating that even strong traditional risk factors for eGFR decline are negated when considering specific biological processes.

The complementary pathway enrichment analysis (section 8.1.6) suggested that many signal transduction reactions were included in the list of proteins showing evidence of an association with eGFR decline, probably because such pathways are core to many functions of the cell and are ubiquitous to many locations in the body. In addition, however, pathways involving extracellular matrix (ECM) organisation and regulation of insulin-like growth factor (IGF) transport and uptake by IGF binding proteins (IGFBPs) were also identified. This finding supports the fact that PTPRS and IGFBP6 did not by chance appear to be associated with CKD progression because of the mere fact that many ECM and IGF regulatory proteins were included in the targeted cardiometabolic Olink® panels. These proteins and biological pathways are further contextualised below.

8.2.1 Receptor-type protein-tyrosine phosphatase Ơ (sigma)

The exact process by which PTPRS may modulate CKD progression is not documented but may be multifactorial. Binding of the trans-membrane PTPRS, an extracellular protein, with its ligand facilitates tyrosine de-phosphorylation which modulates intracellular signalling.(282) Chondroitin sulphate and heparan sulphate proteoglycans (CSPGs and HSPGs) act as ligands.(283) PTPRS is expressed specifically in cartilage, neuronal tissue, haematopoietic stem/progenitor cells and by the vasculature.(284,285)

There is no literature implicating PTPRS in CKD progression but it has been implicated in ulcerative colitis activity(286), joint destruction in rheumatoid arthritis(287), murine embryological urogenital development(288), and was highly expressed in small arteries of people with CKD.(289) In the latter study, normotensive people with normal kidney function and hypertensive people (BP >135/85 mmHg and/or prescribed antihypertensive medication) with or without an eGFR < 60ml/min/1.73m^2 underwent transcriptomic profiling of biopsied subcutaneous small arteries. Up-regulation of PTPRS messenger RNA was demonstrated in those with CKD and hypertension (mean 45±11 ml/min/1.73m^2) especially in vascular smooth muscle cells and adventitial fibroblasts. However, there was only a 5-mmHg difference in mean systolic BP between those with hypertension without CKD and those with CKD. The individuals in the EQUAL study had a much lower average eGFR at baseline and hypertension was not well controlled compared to the cited study but systolic blood pressure was adjusted for in the current analysis. The role of PTPRS in CKD progression may thus be vascular in origin.

PTPRS may also contribute to CKD progression through its function in neuronal tissue. The action of CSPGs binding to PTPRS is to inhibit central and peripheral nervous system axonal

growth whereas HSPGs have the opposite function.(290) In a *ptprs* gene knockout mouse model, CSPGs did not prevent sympathetic re-innervation following cardiac ischaemic-reperfusion induced injury.(291) Coronary artery disease and heart failure, that may be a cause and consequence of cardiac ischaemia, respectively, were highly prevalent in the EQUAL study population. Relevant to the kidneys, renal sympathetic overactivity increases as CKD progresses and through noradrenaline, contributes to kidney fibrosis.(292,293) Although not directly tested elsewhere or corroborated by the current analysis, PTPRS may stimulate kidney fibrosis by its action on the renal sympathetic supply similar to that found in cardiac sympathetic re-innervation.

The ECM has garnered much attention to the extent that a *matrisome*[29] database of its components has been formulated.(294) The present pathway enrichment analysis found that the Olink® proteins studied were enriched for the ECM organisation pathway. PTPRS is involved in this pathway, which is consistent with fibrosis as a possible driver of CKD progression. Against this, Thrombospondin-2, Cadherin-1 and SPARC-like protein 1, proteins similar to those found to be associated with kidney fibrosis in other literature, and Pro-fibrotic bone morphogenetic 6, Transforming growth factor β receptor type 3 and Tissue inhibitor of metalloproteinase-1 protein were not associated with kidney function loss in this analysis.(295–297) The fact that individual proteins, and PTPRS but not other ECM-related proteins, had a more pronounced effect on kidney function than an entire pathway combined using principal component analysis might indicate the opposing processes between upregulated and downregulated proteins within a given pathway, underscore the

[29] The name given to the specific study of proteins of the extracellular matrix proteome.

dysregulation seen in kidney disease or highlight the particularly prominent role that PTPRS might have in CKD progression.

Nevertheless, antifibrotic agents have been shown to limit CKD progression but only modestly, probably because of the myriad ECM organisation proteins that would need to be targeted to be effective.(298) PTPRS may be a novel anti-fibrotic target.

Interestingly, the low-molecular weight heparin, enoxaparin, which is used as an anticoagulant therapeutically has been found to stimulate the PTPRS receptor in vivo; future work should be conducted in humans to determine if this widely used medication limits either sympathetic overactivity or CKD progression.(299)

A word of caution: since circulating PTPRS was measured in plasma in the current study, it is not possible to speculate the concentration of PTPRS in the kidney (in particular the renal vessels and interstitium) or its sympathetic supply, but its abundance in peripheral blood might mean that there is shedding of PTPRS from its normal locations in nervous and vascular tissues and that it may serve as a proxy measure of vascular and interstitial integrity rather than a causative role in CKD progression.

While the related protein tyrosine phosphatase 1-β has a significant role in angiogenesis, and was not included in Olink® panels, PTPRS ostensibly does not.(300–302) Nevertheless, analysed proteins that are involved in angiogenesis were not associated with eGFR decline in this analysis (Tyrosine-protein kinase receptor Tie-1, Angiopoietin-1 receptor, Vascular

endothelial growth factor D, Angiopoietin-1, Angiopoietin-related protein 3, Angiogenin, Cadherin-1, Platelet-derived growth factor subunit B) contrary to other literature.(303)

Protein tyrosine phosphatases, the family to which PTPRS belongs, are abundant in insulin target tissues which suggests regulation of the insulin receptor by PTPs.(304) Previously, *ptprs* single nucleotide polymorphisms (SNPs) were analysed in Swedish individuals with normal and impaired glucose tolerance and type 2 diabetes mellitus.(305) Three SNPs were shown to confer susceptibility to development of type 2 diabetes mellitus. In the current analysis, diabetes mellitus status was controlled for, indicating that elevated PTPRS plasma levels as a potential marker of CKD progression may be independent from the traditional risk conferred by diabetes mellitus.

8.2.2 Insulin-like growth factor binding protein 6

IGF-I and IGF-II promotes the proliferation of fibroblasts, contributing to ECM expansion and kidney fibrosis, previously reviewed by Nahas, Zhu et al. and Liso et al.(306–308) IGFBP6 primarily transports IGF-II thereby directing IGF-II to its target tissues, prolonging its half-life and regulating its activity.(309) In addition to its systemic presence, IGFs and IGFBPs are synthesized by kidney cells and act in an autocrine and paracrine fashion.(310) Increased circulating IGFBP6 levels may therefore represent an attempt to sequester free IGF-II to increase or decrease local kidney targeted activity. It is noteworthy that while IGFBP3, that carries IGF-I, is the most abundant IGFBP in circulation, is expressed by proximal tubular cells(311) and has been found to increase in people with eGFR < 60ml/min/1.73m^2, it was not associated with eGFR decline in this study.(309,312)

While there is no previous evidence that IGFBP6 is deleterious to kidney function, it has been observed to increase in children with CKD(313), gene expression is upregulated in CKD(314), and its levels have been shown to decrease in adults with KFKRT post-transplantation.(315) In the latter study, although serum IGFBP6 levels decreased, levels were still elevated above those with normal kidney function up to 2 weeks post-transplantation. Also, IGFBP6 was detectable in the urine of normal non-anuric participants, at least explaining that serum elevations may in part be due to decreased urinary excretion but since levels do not completely normalise post-transplantation, this suggests increased production as well. IGFBP6 mRNA is highly expressed in fibroblasts within renal blood vessels in the rat kidney, human vascular smooth muscle cells(316), and to a lesser extent, interstitial cells.(317)

The term 'fibrogenic niche' has recently been coined to encapsulate the microenvironment that promotes fibroblast expansion that results in fibrosis.(230,318) Previously, Tenascin-C was found to be a major constituent in the fibrogenic niche in vivo and in vitro murine unilateral ureteric obstruction and ischaemic-reperfusion induced CKD models.(318) In the current study, it was not found to be associated with CKD progression suggesting a less prominent role of TNC in human advanced kidney disease.

The matrisome database (http://www.pepchem.org/matrisomedb) collects data on ECM proteins that have been detected in normal and diseased tissues through a consortium of participating studies.(294) A query for PTPRS was negative but IGFBP6 was described in studies related to the vasculature. The database is, however, primarily of normal, cancerous, vascular and lung and liver tissue types but not kidney interstitial architecture – there is only one glomerular cell study out of the 17 cited.(294,319) In the absence of data for other kidney

cellular compartments and further glomerular cataloguing, it is difficult to postulate that these proteins are at least *not* prominent in glomerular fibrosis.

In support of a glomerular process, on the other hand, is the finding that in immortalised human and murine models, Hale et al. describe the podocyte as a target for IGF-II and a reduction in IGF-II or knockout of the IGF-I receptor (IGF-IR) causes podocyte death.(320) In this regard, I speculate that IGFBP6 may sequester IGF-II, preventing it from attaching to its IGF-IR, leading to podocyte loss. Proteinuria was unfortunately incompletely recorded in the EQUAL study, and so precludes any further work to test the hypothesis that there might be a role of IGFBPs in worsening glomerular proteinuria as a mechanism for CKD progression.

However, in the Canagliflozin Cardiovascular Assessment Study randomised trial of people with type 2 diabetes mellitus and established cardiovascular disease (age > 30 years) or risk factors for cardiovascular disease (age > 50 years) and eGFR > 30 ml/min.1.73m^2, the related IGFBP7 predicted the renal composite outcome of a sustained reduction in eGFR by 40%, KRT initiation, renal or cardiovascular death, or progression to macroalbuminuria (Hazard Ratio 3.80; 95% CI 3.16, 7.60).(321) No difference was seen in those receiving canagliflozin or placebo for the composite.

Although both IGFBP6 and PTPRS seem to be involved in ECM organisation and fibrosis, they seem to act independently of each other. A STRING-db search (conducted on 19 January 2023) did not show a protein-protein interaction between PTPRS and IGFBP6 based on extant evidence.(236)

8.2.3 Ficolin-2

One of the components of the innate immune system is the complement cascade consisting of three routes: classical, alternative and lectin pathways which promote inflammation, opsonisation and forms the cytolytic membrane attack complex.(322) Inflammation has been associated with CKD progression and the glomerulus is susceptible to complement injury.(98,323) Ficolin-2 (FCN2) is an extracellular protein responsible for activating the lectin pathway by acting as a receptor for pathogen-associated molecular patterns.(324) Such activation aims to defend the host by killing the perceived threat, be it self or non-self, and clear injured cells and necrotic debris. Complement mediated kidney disease may, however, result when such activation is dysregulated.

In a Danish study of patients with systemic lupus erythematosus, a complement-mediated autoimmune disease, low FCN2 levels, stratified by its median, predicted the development of lupus nephritis.(325) In that study, high FCN1 levels predicted development of kidney failure, defined as recorded initiation of KRT in the Danish Renal Registry, but FCN2 did not.

Also from Denmark, FCN2 levels dropped substantially after starting haemodialysis although at that point, it is already too late for intrinsic kidney function recovery so haemodialysis would be a poor intervention unless started at a much higher eGFR.(326) Once transplanted, *fcn2* polymorphisms have been shown to increase susceptibility to delayed graft function and acute rejection post deceased donor kidney-transplantation.(327)

The lectin pathway, which FCN2 activates, results in the formation of C3 convertase, an enzyme that splits C3 into C3a and C3b which then effect their immune function through

chemoattraction and phagocytosis.(328) In the current study, the association between FCN2 and kidney function decline was ameliorated after additional adjustment for RAAS-inhibitors and β-blockers.

There is no direct link between RAAS-inhibition and FCN2. Renin, however, has been shown to cleave C3, analogous to C3 convertase, and renin-blockers inhibit renin-induced cleavage.(329) Also, high levels of C3 promoted tubular epithelial to mesenchymal[30] transition leading to a renin-secreting phenotype in hypertensive rats, which would exacerbate the C3 convertase-like renin effect and trigger the RAA system.(330) Renin cleaves angiotensinogen into angiotensin-I (ANG) and angiotensin-converting enzyme cleaves ANG-I into ANG-II. ANG-II is pro-fibrotic.(331) Given this evidence, RAAS-inhibition may prevent fibrosis that results via C3 induced renin-ANG-II. Further research is needed to explain if FCN2 is specifically involved in RAAS. Neither renin nor angiotensin-converting enzyme 2 were associated with eGFR decline in the current study.

Furthermore, C3b activates C5a which recruits further inflammatory cells. In a murine C5 gene knockout model of tubulointerstitial fibrosis, fibrosis was limited in C5 deficient mice compared to wild-type.(332) Eculizumab is a monoclonal antibody that blocks C5 and low molecular weight heparin inhibits the lectin pathway via C1 esterase which bodes well for potential repurposed therapeutics.(333,334) FCN2 has not specifically been linked to *kidney* fibrosis but *lower* levels predicted more severe *liver* fibrosis in people with non-alcoholic fatty liver disease.(335) In the current study, *higher* levels of FCN2 were associated with eGFR

[30] A type of cell present during embryonic development that differentiates into epithelium.

decline which is not consistent with this finding. The contribution of FCN2 in kidney fibrosis, especially in the presence of RAAS-inhibition, remains elusive.

Interestingly, other proteins which were analysed, Mannose-binding protein C, also an activator of the lectin pathway, Complement C1q tumour necrosis factor-related protein 1 and Pentraxin-3 which activates the classical complement pathway, and regulatory proteins CD59 and thrombomodulin, were not associated with eGFR decline.

Glomerular diseases, often complement or immune-complex related, were only the primary renal disease in ~10% of those recruited in the EQUAL study. This finding would suggest an independent role (possibly through fibrosis) for FCN2 in CKD progression beyond that previously described regarding lectin pathway/ primary complement mediated glomerular disease, such as IgA nephropathy.(336,337) Once again, FCN2 activity would need to be measured at the level of the kidney to support a direct pathogenic role.

8.2.4 Additional proteins proposed using multiple imputation modelling

The multiple imputation model identified additional proteins that may be associated with eGFR decline and computed similar slope estimates as the primary analysis for PTPRS, IGFBP6 and FCN2. The additional proteins, barring CST3 an F7, demonstrate very similar extracellular and immunological functions.

Most relevant to those already discussed, EFEMP1 was associated with eGFR decline in a Mendelian randomisation study in participants of the Framingham Heart Offspring cohort which adds to the finding of fibrosis pathways as a cause of CKD progression.(338)

Immunoglobulin light chains, which IGLC2 is a component of, is profibrotic by activating reactive oxygen system pathways in in vivo experiments.(339) NCAM1, a neuronal cell adhesion molecule, may act as an autoantigen in membranous lupus nephritis and has been found expressed in fibrosed human interstitum.(340–342) Also, serum levels of PCOLCE, was found to be higher in people with CKD compared to those with normal kidney function.(343) Haematopoietic stem cells infiltrate diseased kidneys, attracted by interactions with the mast/stem cell growth factor receptor KIT, analysed here. Mast cells were abundant and stem cell factor ('kit-ligand') was highly expressed in fibrosed kidney interstitium in patients with primary and secondary glomerulonephritis.(344) Further, CD59, a protein that regulates the membrane attack complex which is the final product of the complement cascade, was less expressed on human leukocytes in people with CKD to possibly counter complement hyperactivity.(345)

Furthermore, CKD is a pro-coagulable state and several haemostatic components, including F7 that was identified in this analysis, have been found to increase as CKD progresses.(346,347) Finally, CST3 is a well-known biomarker of kidney function and improves the prediction of progression to KF, cardiovascular events and death compared to creatinine based eGFR equations, especially in older people in whom CST3 is less affected by muscle mass.(348,349) Unfortunately CST3 was not assayed longitudinally so could not be incorporated into the eGFR calculation.

8.2.5 Strengths and limitations

Strengths of this study are that a large panel of proteins were characterised in multiple European nationalities with advanced CKD. Not only were individual proteins analysed, but the biological pathways that they represent were also shown to have an association with eGFR decline. Having said that, there may have been other proteins or pathways that were not tested that may be important.(230) This analysis contributes to the current sparse human data up until now predominated by animal studies and almost non-existent literature of these proteins' association with CKD progression in humans.

A limitation of this study, however, was that there was no control group of people with normal kidney function for comparison and the sample population was mainly of European descent. Proteins were only measured at baseline so changes in their concentration could not be modelled serially as was repeated eGFR. Urine and tissues samples would have been complementary and would have allowed more conclusive mechanistic evidence to support a robust relationship but were not available. In addition, data for some proteins were not available and ACR was missing in a large proportion of patients. The complete case sample size was therefore greatly diminished.

Future work should involve externally reproducing this analysis on other groups of people with CKD in terms of geography (non-European), age (only older participants were included in the EQUAL study) and earlier CKD stages to determine the robustness of the associations.

8.3 CONCLUSION

The progression of CKD is biologically complex involving multiple kidney compartments and dysregulated pathways. Receptor-type tyrosine-protein phosphatase S, Insulin-like growth factor-binding protein 6 and Ficolin-2 with established links to kidney fibrosis, vasculopathy, axonal growth and the complement cascade, were associated with more rapid CKD progression.

Further research is needed to determine if these proteins are causally linked to accelerated progression of CKD in older people or a secondary consequence of shedding from their normal locations or a result of decreased kidney excretion. There is potential for these proteins as markers of kidney pathology *in lieu* of histological confirmation of fibrosis and immunological confirmation of complement activation as well as potential therapeutic targets to decelerate CKD progression.

9 CONCLUSIONS, CONTRIBUTIONS, IMPLICATIONS AND FUTURE DIRECTIONS

9.1 INTRODUCTION

Data about people with kidney disease are increasingly being collected via biological samples and clinical information using computerised automated administrative and electronic health records and warehoused in registries and data centres. In this thesis, the clinical epidemiology of acute and chronic kidney diseases was explored using routinely collected laboratory and cohort study data, allowing the detection of *de novo* AKI, A-on-CKD and CKD, and CKD progression. Linked comorbidity and encounter data allowed characterisation of these disorders and ancillary proteomic data permitted an investigation into the biological underpinnings of kidney function decline.

9.2 SUMMARY OF MAIN RESULTS

This thesis used three separate secondary data sources emanating from the United Kingdom, Western Europe and Cape Town, South Africa to achieve its objectives. The individual studies, results and important take-away points are summarised below.

9.2.1 Investigation of the implementation and consistency of the NHS England-mandated AKI detection algorithm

In 2014, NHS England mandated that laboratories, through its laboratory information management system (LIMS), implement an AKI detection algorithm and submit AKI alerts to the UK Renal Registry (UKRR). In this study, alerts and pre- and post-AKI SCr results submitted to the UKRR were used to establish if AKI alerts were consistent with alerts generated using simulated algorithm code in the same set of longitudinal SCr values. This would indirectly and collectively assess the syntax used to code the AKI algorithm, based on guidelines published by 'Think Kidneys', the extent of alert suppression (alerts not reaching the UKRR) and the submission of alert data by laboratories to the UKRR. Agreement was excellent across LIMSs and most individual laboratories, barring a few. In a sub-group analysis, agreement decreased as the baseline SCr increased, causing concern for possible misclassification in people with CKD. A limitation of this study was the exclusive analysis of laboratories with near-complete data, which may not be representative of all laboratories submitting AKI and SCr data to the UKRR.

9.2.2 Characterisation of healthcare clients ascertained with AKI in the City of Cape Town, 2017 – 2021

The Provincial Health Data Centre (PHDC) of the Western Cape, South Africa, is a health information exchange that uses data collected from healthcare clients in the public healthcare system to ascertain, among other conditions, acute and chronic kidney diseases using a bespoke algorithm. PHDC-ascertained AKI was described for the first time and associated demographics, comorbidities and healthcare utilisation information was used to describe those ascertained with the condition. Notably, the burden of AKI on public health services was immense and injury was severe. Although unique and pragmatic, the PHDC AKI detection

algorithm was not comparable to the NHSE algorithm, which is based on consensus KDIGO guideline criteria, necessitating caution when interpreting the findings. Importantly, the PHDC algorithm may have detected worse AKI in females compared to males and it inherently does not permit AKI detection in people with established CKD.

9.2.3 Clinical epidemiology of healthcare clients ascertained with CKD in the City of Cape Town

The PHDC also records healthcare clients with CKD, based on sustained decreases in eGFR over time, and KRT, based on dialysis and transplantation procedure codes, regularity of renal clinic visits and prescription of medications commonly dispensed to people with KF and transplant recipients. This study aimed to describe the number and attributes of individuals with CKD and on long-term KRT. Uncertainty about who was still prevalent with CKD over time was a limitation. Disengagement with healthcare services has implications for the personal clinical care of the individual as well as monitoring the prevalent burden of CKD in the City. The results suggested that KRT was not accurately captured, necessitating future investigation and validation.

9.2.4 Incidence and characteristics of healthcare clients with A-on-CKD in the City of Cape Town, 2017 – 2021

Since the PHDC does not, by design, permit the detection of AKI in individuals with pre-existing CKD, a modified NHSE algorithm was executed in the SCr results of healthcare clients with PHDC-ascertained CKD in the City of Cape Town to determine the incidence and traits of those with A-on-CKD. The number of incident cases differed by sex, possibly explained by how men and women access healthcare, and by incident/ prevalent CKD populations, limited by the availability of longitudinal SCr data in persons with prevalent CKD before the execution of the

algorithm. Modifications also allowed short-term recovery to be assessed. The assessment of short-term recovery was limited by the lack of future SCr testing in people who did survive the AKI.

9.2.5 Cardiometabolic associations with kidney function decline in older adults with advanced CKD

Longitudinal eGFR and baseline cardiometabolic protein data in older people with advanced CKD from the European Quality Study (EQUAL) were used to investigate the association between cardiometabolic proteins and pathways with progression of CKD. In the primary analysis, higher concentrations of three proteins were associated with eGFR decline. These proteins were involved in extracellular matrix and axonal growth pathways and the complement cascade. It is uncertain whether these proteins or pathways *cause* CKD progression or if they simply have accumulated because of impaired kidney excretion. It was not possible to characterise protein expression at the kidney tissue level, also precluding any conclusion about the direct effects of these ubiquitous proteins on kidney function specifically. Fibrosis can be seen as the common distal pathway leading to KF, no matter what the inciting event or injurious, usually dysregulated, processes that accelerated this trajectory to kidney function loss. This analysis in people with advanced kidney disease, may have captured a fibrogenic niche milieu incidentally at a time when the sequelae of kidney disease were already well established i.e., fibrosis.

9.3 Implications for Translational and Population Health

Before prevention and treatment of AKI, CKD and progression to kidney failure can be contemplated, recognition of the occurrence of kidney disease and decline in kidney function is paramount. Algorithms were implemented to detect kidney disease using commonly requested laboratory tests. Thus, the burden of disease by algorithm-detected AKI and CKD were explored. The South African Renal Registry currently only receives data from renal units voluntarily submitting information on people receiving KRT care.(195) The UKRR, in contrast, receives data on people with algorithm-detected AKI, advanced CKD and on KRT.(74)

The research presented in this thesis will be used to strengthen the algorithms used by the PHDC to detect kidney disease and the findings of the burden and characteristics of AKI, CKD and A-on-CKD will be used to advocate for kidney disease screening and prevention in the province. The PHDC is unique in that it could potentially function as a province-wide kidney health surveillance system collecting data on burden of disease, healthcare utilisation and kidney outcomes across primary and hospital levels of care in the public sector.(79) Much work is necessary however before this can be accomplished, including improving the operationalisation of definitions of kidney diseases and ascertainment of people receiving KRT care. Theoretically, quality indicators such as levels of anaemia, dialysis adequacy and mineral and bone disease control can be incorporated. In the future, private dialysis services could be onboarded as well. Also, a kidney function and disease summary will soon be available as a clinician portal as part of the PHDC health information exchange.(80) This will support clinicians in recognising new and worsening kidney disease.

The UKRR, which receives and processes AKI alerts submitted from laboratories across England and reports on AKI epidemiology and outcomes can be reassured by the findings of this research that the alerts that they receive are valid, at least for laboratories that submit complete data and except for a few laboratories. The results of the analysis presented in this thesis could be shared with laboratories and LIMS providers and prompt investigation and audit of how laboratories generate and process alerts generally and specifically in people with CKD. Specifically, steps need to be audited from the code written to generate the alert, how the alert is processed by the LIMS/ laboratory, and whether alerts that should reach the Registry are suppressed by the laboratory. The UKRR, in consultation with laboratories and LIMS providers, could investigate automatic delivery of alert data to the UKRR to overcome issues of incomplete data files being sent or for laboratories currently not submitting at all, to achieve full compliance with the Patient Safety Alert issued by NHS England.(63)

Although intentionally hypothesis generating, the results of the analysis of cardiometabolic protein associations with CKD progression require further study. Nevertheless, immunological and fibrosis pathways as well as interactions with the vascular wall are promising targets of intervention. At the very least, the analysis highlighted important considerations in the assessment of eGFR decline and complex biological systems that should be taken into account when investigating biological mechanisms of CKD progression.

9.4 FUTURE RESEARCH

Further research is required in the detection of kidney disease, collection of kidney disease attributes and finding mechanisms of kidney disease progression.

The uptake of digital administrative registration and encounter data, health records, laboratory and other data permits the automated detection of kidney disease using consensus SCr and eGFR criteria and diagnostic and procedure codes. However, as demonstrated in this thesis, routinely collected data are complex and clinical and contextual information is often unavailable. Rule-based algorithms are efficient and easy to implement, but several aspects of the continuum from choosing and operationalising the baseline SCr, thresholds used to define kidney disease and misclassification of kidney disease are factors that need attention to ensure correct capture of kidney disease.

There is much work that should still be undertaken in the field of automated AKI detection especially related to the NHS England algorithm and the laboratories in which the algorithm is operational. Since the baseline SCr is required to define and stage AKI, it is imperative to use the most clinically relevant value that generates alerts with the highest sensitivity and specificity.(20,350,351) Further research should be conducted to determine the most appropriate baseline SCr when known, particularly in people with pre-existing CKD and especially when routinely collected data are studied. The most appropriate baseline is likely the value that is most predictive of mortality and adverse kidney outcomes.(352) The possibility of alert suppression should be audited and guidelines should incorporate specific advice. This will ensure that reports of AKI clinical epidemiology generated by the UKRR are certain that all AKI is being captured, especially in people with pre-existing CKD.

Investigation of different thresholds of change (beyond the standard relative and absolute changes proposed by KDIGO), including individualised prediction as exemplified by the precision medicine movement and personalised SCr reference ranges, might show better performance especially in the context of CKD.(353–355) Also, given that laboratories may have operationalised the NHS England algorithm differently, it is pertinent to elucidate how the Think Kidneys best practice guideline has been interpretated and to audit the algorithm syntax that has been coded within LIMSs, especially in laboratories found in this analysis to have poor agreement with the central code.

Limitations of the proteomic analysis of kidney function decline were that the study was restricted to older people with advanced CKD and only serum proteins were assayed. If the aim is to discover proteins that cause kidney function decline, then analysis of people of all ages with normal and all stages of CKD should be studied. Also, the serum, urine and tissue compartments (kidney biopsy) should directly be tested at multiple time points. Kidney function should be measured and not estimated using the imperfect eGFR and proteinuria should be quantified longitudinally. Individual proteins as well as metabolites, genomic and epigenetic signatures, and core biological pathways should be investigated. This would constitute the ideal study to provide robust conclusions about the causes of kidney function decline, albeit expensive and logistically difficult, and permit the discovery of therapeutic interventions that would impede CKD development and progression. This type of investigation could be incorporated into future trials of candidate drugs to slow CKD progression.

9.5 CLOSING REMARKS

Understanding kidney disease epidemiology starts with its recognition and identifying biological mechanisms of its progression. Rule-based algorithms are useful to monitor the occurrence of new kidney disease. The NHSE algorithm, a modified NHSE algorithm and a kidney disease algorithm used by the PHDC were used in this thesis to describe the clinical epidemiology of acute and chronic kidney diseases. Algorithms should be based on well recognised criteria so that kidney disease detection is uniform. Kidney disease episodes that emanate from different laboratories or that are based on information collected at different facilities or from multiple source systems need to be validated and expected to be consistent across diverse settings and levels of kidney dysfunction. Once kidney disease has been established, novel proteomics and bioinformatic tools show promise to appreciate complex biological systems and pathways that may explain kidney disease progression, with the ultimate goal of preventing kidney failure and premature death, and improving the quality of life of people living with kidney disease.

Administrative data (the digitome) and the proteome provide unique opportunities to detect and explain kidney disease. However, limitations of the methods used to scrutinise the digitome and proteome, such as kidney disease misclassification, the complex way healthcare clients access and engage with medical facilities, how their data are collated to infer disease episodes, the way in which disease episodes are defined, incomplete or missing data and the inability to derive causal biological relationships, were identified for further refinement and investigation in the future.

10 REFERENCE LIST

1. Nielsen S, Kwon TH, Fenton RA, Prætorious J. Anatomy of the Kidney. In: Brenner and Rector's The Kidney. 2012.

2. Lote CJ. Principles of renal physiology. Principles of Renal Physiology. 2013.

3. Perrone RD, Madias NE, Levey AS. Serum creatinine as an index of renal function: New insights into old concepts. Clin Chem. 1992;

4. Levey AS, Coresh J, Greene T, Marsh J, Stevens LA, Kusek JW, et al. Expressing the modification of diet in renal disease study equation for estimating glomerular filtration rate with standardized serum creatinine values. Clin Chem. 2007;

5. Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;

6. Delgado C, Baweja M, Crews DC, Eneanya ND, Gadegbeku CA, Inker LA, et al. A Unifying Approach for GFR Estimation: Recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease. Am J Kidney Dis Off J Natl Kidney Found. 2022 Feb;79(2):268-288.e1.

7. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. N Engl J Med. 2012;

8. Shlipak MG, Matsushita K, Ärnlöv J, Inker LA, Katz R, Polkinghorne KR, et al. Cystatin C versus creatinine in determining risk based on kidney function. N Engl J Med. 2013 Sep 5;369(10):932–43.

9. Schaeffner ES, Ebert N, Delanaye P, Frei U, Gaedeke J, Jakob O, et al. Two novel equations to estimate kidney function in persons aged 70 years or older. Ann Intern Med. 2012;

10. Chawla LS, Bellomo R, Bihorac A, Goldstein SL, Siew ED, Bagshaw SM, et al. Acute kidney disease and renal recovery: Consensus report of the Acute Disease Quality Initiative (ADQI) 16 Workgroup. Nat Rev Nephrol. 2017;

11. Kellum JA, Lameire N, Aspelin P, Barsoum RS, Burdmann EA, Goldstein SL, et al. Kidney disease: Improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury. Kidney Int Suppl. 2012;2(1):1–138.

12. Bellomo R, Ronco C, Kellum JA, Mehta RL, Palevsky P. Acute renal failure - definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. In: Critical care (London, England). 2004.

13. Mehta RL, Kellum JA, Shah SV, Molitoris BA, Ronco C, Warnock DG, et al. Acute kidney injury network: Report of an initiative to improve outcomes in acute kidney injury. Crit Care. 2007;

14. Mehta RL, Cerdá J, Burdmann EA, Tonelli M, García-García G, Jha V, et al. International Society of Nephrology's 0by25 initiative for acute kidney injury (zero preventable deaths by 2025): A human rights case for nephrology. The Lancet. 2015;

15. Waikar SS, Betensky RA, Emerson SC, Bonventre JV. Imperfect gold standards for kidney injury biomarker evaluation. J Am Soc Nephrol. 2012;23(1):13–21.

16. M. TJ, Jennifer C, et al. Potential Effects of Aggressive Decongestion During the Treatment of Decompensated Heart Failure on Renal Function and Survival. Circulation. 2010 Jul 20;122(3):265–72.

17. Chu R, Li C, Wang S, Zou W, Liu G, Yang. L. Assessment of KDIGO definitions in patients with histopathologic evidence of acute renal disease. Clin J Am Soc Nephrol. 2014;9(7):1175–82.

18. Haase M, Bellomo R, Devarajan P, Schlattmann P, Haase-Fielitz A, Bagshaw SM, et al. Accuracy of Neutrophil Gelatinase-Associated Lipocalin (NGAL) in Diagnosis and Prognosis in Acute Kidney Injury: A Systematic Review and Meta-analysis. Am J Kidney Dis. 2009;

19. Waikar SS, Bonventre JV. Creatinine Kinetics and the Definition of Acute Kidney Injury. J Am Soc Nephrol JASN. 2009 Mar;20(3):672–9.

20. Siew ED, Alp Ikizler T, Matheny ME, Shi Y, Schildcrout JS, Danciu I, et al. Estimating baseline kidney function in hospitalized patients with impaired kidney function. Clin J Am Soc Nephrol. 2012;7(5):712–9.

21. Bernier-Jean A, Beaubien-Souligny W, Goupil R, Madore F, Paquette F, Troyanov S, et al. Diagnosis and outcomes of acute kidney injury using surrogate and imputation methods for missing preadmission creatinine values. BMC Nephrol. 2017;18(1):1–9.

22. Siew ED, Matheny ME. Choice of reference serum creatinine in defining acute kidney injury. Nephron. 2015;131(2):107–12.

23. Kellum JA. How can we define recovery after acute kidney injury? Considerations from epidemiology and clinical trial design. Nephron - Clin Pract. 2014;127(1–4):81–8.

24. Sharma A, Mucino MJ, Ronco C. Renal functional reserve and renal recovery after acute kidney injury. In: Nephron - Clinical Practice. 2014.

25. Mehta S, Chauhan K, Patel A, Patel S, Pinotti R, Nadkarni GN, et al. The prognostic importance of duration of AKI: A systematic review and meta-analysis. BMC Nephrol. 2018;

26. Levin A, Stevens PE, Bilous RW, Coresh J, De Francisco ALM, De Jong PE, et al. Kidney disease: Improving global outcomes (KDIGO) CKD work group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl. 2013;

27. Neuen BL, Weldegiorgis M, Herrington WG, Ohkuma T, Smith M, Woodward M. Changes in GFR and Albuminuria in Routine Clinical Practice and the Risk of Kidney Disease Progression. Am J Kidney Dis. 2021 Sep 1;78(3):350-360.e1.

28. Chu CD, Xia F, Du Y, Singh R, Tuot DS, Lamprea-Montealegre JA, et al. Estimated Prevalence and Testing for Albuminuria in US Adults at Risk for Chronic Kidney Disease. JAMA Netw Open. 2023 Jul 27;6(7):e2326230.

29. Chu CD, Powe NR, Shlipak MG, Scherzer R, Tummalapalli SL, Estrella MM, et al. Albuminuria testing and nephrology care among insured US adults with chronic kidney disease: a missed opportunity. BMC Prim Care. 2022 Nov 24;23(1):299.

30. Fraser SD, Parkes J, Culliford D, Santer M, Roderick PJ. Timeliness in chronic kidney disease and albuminuria identification: a retrospective cohort study. BMC Fam Pract. 2015 Feb 13;16(1):18.

31. Lindeman RD, Tobin J, Shock NW. Longitudinal Studies on the Rate of Decline in Renal Function with Age. J Am Geriatr Soc. 1985;

32. Orlandi PF, Xie D, Yang CW, Cohen J. Slope of Kidney Function and its Association with Longitudinal Mortality and Cardiovascular Disease among Individuals with CKD. J Am Soc Nephrol. 2020;

33. Abeysekera RA, Healy HG, Wang Z, Cameron AL, Hoy WE. Heterogeneity in patterns of progression of chronic kidney disease. Intern Med J. 2021 Feb;51(2):220–8.

34. O'Hare AM, Batten A, Burrows NR, Pavkov ME, Taylor L, Gupta I, et al. Trajectories of Kidney Function Decline in the 2 Years Before Initiation of Long-term Dialysis. Am J Kidney Dis. 2012 Apr 1;59(4):513–22.

35. Levey AS, Eckardt KU, Dorman NM, Christiansen SL, Hoorn EJ, Ingelfinger JR, et al. Nomenclature for kidney function and disease: report of a Kidney Disease: Improving Global Outcomes (KDIGO) Consensus Conference. Kidney Int. 2020 Jun 1;97(6):1117–29.

36. Brück K, Jager KJ, Zoccali C, Bello AK, Minutolo R, Ioannou K, et al. Different rates of progression and mortality in patients with chronic kidney disease at outpatient nephrology clinics across Europe. Kidney Int. 2018 Jun;93(6):1432–41.

37. Zeng X, McMahon GM, Brunelli SM, Bates DW, Waikar SS. Incidence, outcomes, and comparisons across definitions of AKI in hospitalized individuals. Clin J Am Soc Nephrol. 2014;9(1):12–20.

38. Susantitaphong P, Cruz DN, Cerda J, Abulfaraj M, Alqahtani F, Koulouridis I, et al. World incidence of AKI: A meta-analysis. Clin J Am Soc Nephrol. 2013;8(9):1482–93.

39. Campbell CA, Li L, Kotwal S, Georgiou A, Horvath AR, Westbrook J, et al. Under-detection of Acute Kidney Injury in Hospitalised Patients – A Retrospective, multi-site, longitudinal study. Intern Med J [Internet]. 2019 Feb 28;n/a(n/a). Available from: https://doi.org/10.1111/imj.14264

40. Mehta RL, Burdmann EA, Cerdá J, Feehally J, Finkelstein F, García-García G, et al. Recognition and management of acute kidney injury in the International Society of Nephrology 0by25 Global Snapshot: A multinational cross-sectional study. The Lancet. 2016;387(10032):2017–25.

41. Adu D, Okyere P, Boima V, Matekole M, Osafo C. Community-acquired acute kidney injury in adults in Africa. Clin Nephrol. 2016;

42. Dlamini TAL, Heering PJ, Chivese T, Rayner B. A prospective study of the demographics, management and outcome of patients with acute kidney injury in Cape Town, South Africa. PLoS ONE. 2017;

43. James MT, Grams ME, Woodward M, Elley CR, Green JA, Wheeler DC, et al. A meta-analysis of the association of estimated GFR, albuminuria, diabetes mellitus, and hypertension with acute kidney injury. Am J Kidney Dis. 2015;

44. Neugarten J, Golestaneh L. Female sex reduces the risk of hospital-associated acute kidney injury: A meta-analysis. BMC Nephrol. 2018;

45. Gameiro J, Agapito Fonseca J, Jorge S, Lopes JA. Acute kidney injury in HIV-infected patients: a critical review. HIV Med. 2019;

46. Khalili H, Bairami S, Kargar M. Antibiotics induced acute kidney injury: Incidence, risk factors, onset time and outcome. Acta Med Iran. 2013;

47. Ungprasert P, Cheungpasitporn W, Crowson CS, Matteson EL. Individual non-steroidal anti-inflammatory drugs and risk of acute kidney injury: A systematic review and meta-analysis of observational studies. Eur J Intern Med. 2015;

48. Yang Y, George KC, Shang WF, Zeng R, Ge SW, Xu G. Proton-pump inhibitors use, and risk of acute kidney injury: A meta-analysis of observational studies. Drug Des Devel Ther. 2017;

49. McDonald JS, McDonald RJ, Comin J, Williamson EE, Katzberg RW, Murad MH, et al. Frequency of acute kidney injury following intravenous contrast medium administration: A systematic review and meta-analysis. Radiology. 2013;

50. Naicker S, Aboud O, Gharbi MB. Epidemiology of Acute Kidney Injury in Africa. Semin Nephrol. 2008;

51. Coca SG, Singanamala S, Parikh CR. Chronic kidney disease after acute kidney injury: A systematic review and meta-analysis. Kidney Int. 2012;81(5):442–8.

52. Sawhney S, Marks A, Fluck N, Levin A, Prescott G, Black C. Intermediate and Long-term Outcomes of Survivors of Acute Kidney Injury Episodes: A Large Population-Based Cohort Study. Am J Kidney Dis. 2017;

53. NCEPOD - Acute Kidney Injury: Adding Insult to Injury Report (2009) [Internet]. [cited 2022 Jun 4]. Available from: https://www.ncepod.org.uk/2009aki.html

54. Grams ME, Waikar SS, MacMahon B, Whelton S, Ballew SH, Coresh J. Performance and limitations of administrative data in the identification of AKI. Clin J Am Soc Nephrol. 2014;9(4):682–9.

55. Logan R, Davey P, De Souza N, Baird D, Guthrie B, Bell S. Assessing the accuracy of ICD-10 coding for measuring rates of and mortality from acute kidney injury and the impact of electronic alerts: an observational cohort study. Clin Kidney J. 2019;1–8.

56. Wilson FP, et al. Electronic health record alerts for acute kidney injury: multicenter, randomized clinical trial. BMJ [Internet]. 2021 Jan 18 [cited 2022 Jun 7];372. Available from: https://pubmed.ncbi.nlm.nih.gov/33461986/

57. Flynn N, Dawnay A. A simple electronic alert for acute kidney injury. Ann Clin Biochem. 2015;52(2):206–12.

58. Baird D, De Souza N, Logan R, Walker H, Guthrie B, Bell S. Impact of electronic alerts for acute kidney injury on patient outcomes: interrupted time-series analysis of population cohort data. Clin Kidney J. 2021 Feb 16;14(2):639–46.

59. Fenna K, Erasmus RT, Zemlin AE. Hospital-acquired acute kidney injury prevalence in in adults at a south african tertiary hospital. Afr Health Sci. 2019;19(2):2189–97.

60. Holmes J, Geen J, Williams JD, Phillips AO. Recurrent acute kidney injury: predictors and impact in a large population-based cohort. Nephrol Dial Transplant. 2020 Aug 1;35(8):1361–9.

61. Koeze J, Keus F, Dieperink W, Van der Horst ICC, Zijlstra JG, Van Meurs M. Incidence, timing and outcome of AKI in critically ill patients varies with the definition used and the addition of urine output criteria. BMC Nephrol. 2017;18(1):1–9.

62. Kellum JA, Sileanu FE, Murugan R, Lucko N, Shaw AD, Clermont G. Classifying AKI by urine output versus serum creatinine level. J Am Soc Nephrol. 2015;26(9):2231–8.

63. Selby NM, Hill R, Fluck RJ. Standardizing the Early Identification of Acute Kidney Injury: The NHS England National Patient Safety Alert. Nephron. 2015;131(2):113–7.

64. Santosh Tirunagari, S Bull, F Christopher. Automatically Detecting AKI Events from Primary Care Records. In Nottingham, United Kingdom; 2017. Available from: https://openresearch.surrey.ac.uk/esploro/outputs/conferencePresentation/Automatically-Detecting-AKI-Events-from-Primary/99512129702346

65. Porter CJ, Juurlink I, Bisset LH, Bavakunji R, Mehta RL, Devonald MAJ. A real-time electronic alert to improve detection of acute kidney injury in a large teaching hospital. Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc. 2014;

66. Sakhuja A, Bataineh A, Dealmeida D, Bilderback A, Ambrosino R, Fuhrman DY, et al. Creating a High-Specificity Acute Kidney Injury Detection System for Clinical and Research Applications. Am J Kidney Dis. 2021 Nov;78(5):764–6.

67. Sawhney S, Fluck N, Marks A, Prescott G, Simpson W, Tomlinson L, et al. Acute kidney injury - How does automated detection perform? Nephrol Dial Transplant. 2015;30(11):1853–61.

68. Sawhney S, Marks A, Ali T, Clark L, Fluck N, Prescott GJ, et al. Maximising Acute Kidney Injury Alerts – A Cross-Sectional Comparison with the Clinical Diagnosis. PLOS ONE. 2015 Jun 30;10(6):e0131909.

69. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical Decision Support for In-Hospital AKI. J Am Soc Nephrol. 2018;29(2):654–60.

70. Selby NM, Crowley L, Fluck RJ, McIntyre CW, Monaghan J, Lawson N, et al. Use of electronic results reporting to diagnose and monitor AKI in hospitalized patients. Clin J Am Soc Nephrol. 2012;

71. Ahmed A, Vairavan S, Akhoundi A, Wilson G, Chiofolo C, Chbat N, et al. Development and validation of electronic surveillance tool for acute kidney injury: A retrospective analysis. J Crit Care. 2015;30(5):988–93.

72. UK Kidney Association. UK Renal Registry Acute Kidney Injury in England. 2020.

73. Murphy E, Burns A, Murtagh FEM, Rooshenas L, Caskey FJ. The Prepare for Kidney Care Study: prepare for renal dialysis versus responsive management in advanced chronic kidney disease. Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc. 2021 May 27;36(6):975–82.

74. UK Renal Registry (2021) UK Renal Registry 23rd Annual Report – data to 31/12/2019, Bristol, UK. Available from renal.org/audit-research/annual-report.

75. Maluleke R. Stats in Brief, Statisitcs South Africa annual report [Internet]. 2019. Available from: statssa.gov.za

76. Western Cape Government annual performance plan 2019 - 2020 [Internet]. Cape Town; 2020. Available from: westerncape.gov.za

77. City of Cape Town Integrated annual report 2017 - 2018 [Internet]. Cape Town; 2018. Available from: capetown.gov.za

78. Western Cape Government annual performance plan 2018 - 2019 [Internet]. Cape Town; 2019 p. 1–197. Available from: westerncape.gov.za

79. Boulle A, Heekes A, et al. International Journal of Data Centre Profile : The Provincial Health Data Centre of the Western Cape. Int J Popul Data Sci. 2019;4(October).

80. Foster R, Heekes A, Hussey H, Smith M, Mutemaringa T, Phelanyane F, et al. Beyond Interoperability with The Single Patient Viewer: A Clinical Portal to Access Integrated Patient Records. Int J Popul Data Sci [Internet]. 2020 Dec 7 [cited 2022 Jun 30];5(5). Available from: https://ijpds.org/article/view/1614

81. Brück K, Jager KJ, Dounousi E, Kainz A, Nitsch D, Ärnlöv J, et al. Methodology used in studies reporting chronic kidney disease prevalence: A systematic literature review. Nephrol Dial Transplant. 2015;

82. Bikbov B, Purcell CA, Levey AS, Smith M, Abdoli A, Abebe M, 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;

83. United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Ageing 2019: Highlights (ST/ESA/SER.A/430).

84. Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, Lasserson DS, et al. Global prevalence of chronic kidney disease - A systematic review and meta-analysis. PLoS ONE. 2016;

85. Kaze AD, Ilori T, Jaar BG, Echouffo-Tcheugui JB. Burden of chronic kidney disease on the African continent: A systematic review and meta-analysis. BMC Nephrol. 2018;

86. Fabian J, Kalyesubula R, Mkandawire J, Hansen CH, Nitsch D, Musenge E, et al. Measurement of kidney function in Malawi, South Africa, and Uganda: a multicentre cohort study. Lancet Glob Health. 2022 Aug;10(8):e1159–69.

87. Osman MA, Alrukhaimi M, Ashuntantang GE, Bellorin-Font E, Benghanem Gharbi M, Braam B, et al. Global nephrology workforce: gaps and opportunities toward a sustainable kidney care system. Kidney Int Suppl. 2018;8(2):52–63.

88. Jones E, Rayner BL, Effa EE, Okpechi IG, Schmitz M, Heering PJ. Survey on available treatment for acute kidney injury in the Southern African Development Community and Nigeria: Are we ready for zero deaths by 2025 in sub-Saharan Africa? BMJ Open. 2019;9(8):1–7.

89. Yeh HC, Ting IW, Huang HC, Chiang HY, Kuo CC. Acute Kidney Injury in the Outpatient Setting Associates with Risk of End-Stage Renal Disease and Death in Patients with CKD. Sci Rep. 2019 Dec;9(1):17658.

90. Tsai WC, Wu HY, Peng YS, Ko MJ, Wu MS, Hung KY, et al. Risk Factors for Development and Progression of Chronic Kidney Disease. Medicine (Baltimore). 2016;

91. Hsu CY, Chinchilli VM, Coca S, Devarajan P, Ghahramani N, Go AS, et al. Post-Acute Kidney Injury Proteinuria and Subsequent Kidney Disease Progression: The Assessment, Serial Evaluation, and Subsequent Sequelae in Acute Kidney Injury (ASSESS-AKI) Study. JAMA Intern Med. 2020;94143:1–9.

92. Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al. A Predictive Model for Progression of Chronic Kidney Disease to Kidney Failure. JAMA. 2011 Apr 20;305(15):1553–9.

93. Evans M, Grams ME, Sang Y, Astor BC, Blankestijn PJ, Brunskill NJ, et al. Risk Factors for Prognosis in Patients With Severely Decreased GFR. Kidney Int Rep. 2018 May;3(3):625–37.

94. O'Hare AM, Choi AI, Bertenthal D, Bacchetti P, Garg AX, Kaufman JS, et al. Age Affects Outcomes in Chronic Kidney Disease. J Am Soc Nephrol. 2007 Oct 1;18(10):2758–65.

95. Wongrakpanich S, Susantitaphong P, Isaranuwatchai S, Chenbhanich J, Eiam-Ong S, Jaber BL. Dialysis Therapy and Conservative Management of Advanced Chronic Kidney Disease in the Elderly: A Systematic Review. Nephron. 2017;137(3):178–89.

96. Thomas B, Matsushita K, Abate KH, Al-Aly Z, Ärnlöv J, Asayama K, et al. Global cardiovascular and renal outcomes of reduced GFR. J Am Soc Nephrol. 2017;28(7):2167–79.

97. Naimark DMJ, Grams ME, Matsushita K, Black C, Drion I, Fox CS, et al. Past Decline Versus Current eGFR and Subsequent Mortality Risk. J Am Soc Nephrol. 2016 Aug 1;27(8):2456–66.

98. Amdur RL, Feldman HI, Gupta J, Yang W, Kanetsky P, Shlipak M, et al. Inflammation and Progression of CKD: The CRIC Study. Clin J Am Soc Nephrol. 2016 Sep 7;11(9):1546–56.

99. Critselis E, Lambers Heerspink H. Utility of the CKD273 peptide classifier in predicting chronic kidney disease progression. Nephrol Dial Transplant. 2016 Feb 1;31(2):249–54.

100. Abbate M, Zoja C, Remuzzi G. How Does Proteinuria Cause Progressive Renal Damage? J Am Soc Nephrol. 2006 Nov 1;17(11):2974–84.

101. Vree TB, Guelen PJ, Jongman-Nix B, Walenkamp GH. The relationship between the renal clearance of creatinine and the apparent renal clearance of beta-2-microglobulin in patients with normal and impaired kidney function. Clin Chim Acta Int J Clin Chem. 1981 Jul 18;114(1):93–9.

102. Grams ME, Surapaneni A, Chen J, Zhou L, Yu Z, Dutta D, et al. Proteins Associated with Risk of Kidney Function Decline in the General Population. J Am Soc Nephrol. 2021 Sep 1;32(9):2291–302.

103. Carlsson AC, Ingelsson E, Sundström J, Carrero JJ, Gustafsson S, Feldreich T, et al. Use of proteomics to investigate kidney function decline over 5 years. Clin J Am Soc Nephrol. 2017;

104. Mackensen-Haen S, Bader R, Grund KE, Bohle A. Correlations between renal cortical interstitial fibrosis, atrophy of the proximal tubules and impairment of the glomerular filtration rate. Clin Nephrol. 1981 Apr;15(4):167–71.

105. Brenner BM, Mackenzie HS. Nephron endowment and the pathogenesis of chronic renal failure. In: Touraine JL, Traeger J, Bétuel H, Dubernard JM, Revillard JP, Dupuy C, editors. Late Graft Loss: Proceedings of the 28th Conference on Transplantation and Clinical Immunology, 3–5 June, 1996 [Internet]. Dordrecht: Springer Netherlands; 1997 [cited 2022 Jun 3]. p. 93–100. (Transplantation and Clinical Immunology). Available from: https://doi.org/10.1007/978-94-011-5434-5_12

106. Kriz W, Gretz N, Lemley KV. Progression of glomerular diseases: Is the podocyte the culprit? Kidney Int. 1998 Sep;54(3):687–97.

107. Jotwani V, Garimella PS, Katz R, Malhotra R, Bates J, Cheung AK, et al. Tubular Biomarkers and Chronic Kidney Disease Progression in SPRINT Participants. Am J Nephrol. 2020;51(10):797–805.

108. Amdur RL, Feldman HI, Dominic EA, Anderson AH, Beddhu S, Rahman M, et al. Use of Measures of Inflammation and Kidney Function for Prediction of Atherosclerotic Vascular Disease Events and Death in Patients With CKD: Findings From the CRIC Study. Am J Kidney Dis Off J Natl Kidney Found. 2019 Mar;73(3):344–53.

109. Cheung KL, Zakai NA, Callas PW, Howard G, Mahmoodi BK, Peralta CA, et al. Mechanisms and mitigating factors for venous thromboembolism in chronic kidney disease: the REGARDS study. J Thromb Haemost. 2018;16(9):1743–52.

110. Wattanakit K, Cushman M, Stehman-Breen C, Heckbert SR, Folsom AR. Chronic kidney disease increases risk for venous thromboembolism. J Am Soc Nephrol JASN. 2008 Jan;19(1):135–40.

111. Rabelink TJ, Zwaginga JJ, Koomans HA, Sixma JJ. Thrombosis and hemostasis in renal disease. Kidney Int. 1994 Aug;46(2):287–96.

112. Yang K, Du C, Wang X, Li F, Xu Y, Wang S, et al. Indoxyl sulfate induces platelet hyperactivity and contributes to chronic kidney disease–associated thrombosis in mice. Blood. 2017 May 11;129(19):2667–79.

113. Sedaghat S, Vries PS de, Boender J, Sonneveld MAH, Hoorn EJ, Hofman A, et al. von Willebrand Factor, ADAMTS13 Activity, and Decline in Kidney Function: A Population-Based Cohort Study. Am J Kidney Dis. 2016 Nov 1;68(5):726–32.

114. Fried L, Solomon C, Shlipak M, Seliger S, Stehman-Breen C, Bleyer AJ, et al. Inflammatory and Prothrombotic Markers and the Progression of Renal Disease in Elderly Individuals. J Am Soc Nephrol. 2004 Dec 1;15(12):3184–91.

115. Witko-Sarsat V, Friedlander M, Capeillère-Blandin C, Nguyen-Khoa T, Nguyen AT, Zingraff J, et al. Advanced oxidation protein products as a novel marker of oxidative stress in uremia. Kidney Int. 1996 May;49(5):1304–13.

116. Zhou LL, Cao W, Xie C, Tian J, Zhou Z, Zhou Q, et al. The receptor of advanced glycation end products plays a central role in advanced oxidation protein products-induced podocyte apoptosis. Kidney Int. 2012 Oct;82(7):759–70.

117. Jafar TH, Stark PC, Schmid CH, Landa M, Maschio G, de Jong PE, et al. Progression of chronic kidney disease: the role of blood pressure control, proteinuria, and angiotensin-converting enzyme inhibition: a patient-level meta-analysis. Ann Intern Med. 2003 Aug 19;139(4):244–52.

118. Mahmoodi BK, Gansevoort RT, Veeger NJGM, Matthews AG, Navis G, Hillege HL, et al. Microalbuminuria and risk of venous thromboembolism. JAMA. 2009 May 6;301(17):1790–7.

119. Matsushita K, et al. Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. Lancet Diabetes Endocrinol [Internet]. 2015 Jul [cited 2022 May 17];3(7). Available from: https://pubmed.ncbi.nlm.nih.gov/26028594/

120. Salmon AHJ, Satchell SC. Endothelial glycocalyx dysfunction in disease: albuminuria and increased microvascular permeability. J Pathol. 2012 Mar;226(4):562–74.

121. Martens RJH, Houben AJHM, Kooman JP, Berendschot TTJM, Dagnelie PC, van der Kallen CJH, et al. Microvascular endothelial dysfunction is associated with albuminuria: the Maastricht Study. J Hypertens. 2018 May;36(5):1178–87.

122. Incalza MA, D'Oria R, Natalicchio A, Perrini S, Laviola L, Giorgino F. Oxidative stress and reactive oxygen species in endothelial dysfunction associated with cardiovascular and metabolic diseases. Vascul Pharmacol. 2018 Jan;100:1–19.

123. Briet M, Collin C, Karras A, Laurent S, Bozec E, Jacquot C, et al. Arterial Remodeling Associates with CKD Progression. J Am Soc Nephrol. 2011 May 1;22(5):967–74.

124. deFilippi CR, Herzog CA. Interpreting Cardiac Biomarkers in the Setting of Chronic Kidney Disease. Clin Chem. 2017 Jan 1;63(1):59–65.

125. Bagshaw SM, Cruz DN, Aspromonte N, Daliento L, Ronco F, Sheinfeld G, et al. Epidemiology of cardio-renal syndromes: workgroup statements from the 7th ADQI Consensus Conference. Nephrol Dial Transplant. 2010 May 1;25(5):1406–16.

126. Schefold JC, Filippatos G, Hasenfuss G, Anker SD, von Haehling S. Heart failure and kidney dysfunction: epidemiology, mechanisms and management. Nat Rev Nephrol. 2016 Oct;12(10):610–23.

127. Koukoulaki M, Papachristou E, Kalogeropoulou C, Papathanasiou M, Zampakis P, Vardoulaki M, et al. Increased Prevalence and Severity of Coronary Artery Calcification in Patients with Chronic Kidney Disease Stage III and IV. Nephron Extra. 2012 Jan;2(1):192–204.

128. Nakamura S, Ishibashi-Ueda H, Niizuma S, Yoshihara F, Horio T, Kawano Y. Coronary Calcification in Patients with Chronic Kidney Disease and Coronary Artery Disease. Clin J Am Soc Nephrol. 2009 Dec 1;4(12):1892–900.

129. Bansal N, Zelnick L, Shlipak MG, Anderson A, Christenson R, Deo R, et al. Cardiac and Stress Biomarkers and Chronic Kidney Disease Progression: The CRIC Study. Clin Chem. 2019 Nov 1;65(11):1448–57.

130. Kim Y, Matsushita K, Sang Y, Grams ME, Skali H, Shah AM, et al. Association of high-sensitivity cardiac troponin T and natriuretic peptide with incident ESRD: the Atherosclerosis Risk in Communities (ARIC) study. Am J Kidney Dis Off J Natl Kidney Found. 2015 Apr;65(4):550–8.

131. Wang K, Zelnick LR, Anderson A, Cohen J, Dobre M, Deo R, et al. Cardiac Biomarkers and Risk of Mortality in CKD (the CRIC Study). Kidney Int Rep. 2020 Nov;5(11):2002–12.

132. Bullen AL, Katz R, Jotwani V, Garimella PS, Lee AK, Estrella MM, et al. Biomarkers of Kidney Tubule Health, CKD Progression, and Acute Kidney Injury in SPRINT (Systolic Blood Pressure Intervention Trial) Participants. Am J Kidney Dis. 2021;78(3):361-368.e1.

133. The Gene Ontology Consortium, Carbon S, Douglass E, Good BM, Unni DR, Harris NL, et al. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021 Jan 8;49(D1):D325–34.

134. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000 May;25(1):25–9.

135. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003 Nov 1;13(11):2498–504.

136. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009 Jan;4(1):44–57.

137. Joshi-Tope G. Reactome: a knowledgebase of biological pathways. Nucleic Acids Res. 2004 Dec 17;33(Database issue):D428–32.

138. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics. 2009 Nov 15;25(22):3045–6.

139. Mehta R, Cai X, Lee J, Xie D, Wang X, Scialla J, et al. Serial Fibroblast Growth Factor 23 Measurements and Risk of Requirement for Kidney Replacement Therapy: The CRIC (Chronic Renal Insufficiency Cohort) Study. Am J Kidney Dis. 2020;

140. Jager KJ, Ocak G, Drechsler C, Caskey FJ, Evans M, Postorino M, et al. The EQUAL study: A European study in chronic kidney disease stage 4 patients. Nephrol Dial Transplant. 2012;27(SUPPL. 3):27–31.

141. Assarsson E, Lundberg M, Holmquist G, Björkesten J, Thorsen SB, Ekman D, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS ONE. 2014;9(4).

142. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. 1960;

143. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008;

144. Gwet KL. Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters. Advanced Analytics, LLC; 2014.

145. Klein D. Implementing a General Framework for Assessing Interrater Agreement in Stata. Stata J Promot Commun Stat Stata. 2018 Dec;18(4):871–901.

146. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159–74.

147. Sawhney S, Robinson HA, Veer SN van der, Hounkpatin HO, Scale TM, Chess JA, et al. Acute kidney injury in the UK: a replication cohort study of the variation across three regional populations. BMJ Open. 2018 Jun 1;8(6):e019435.

148. Kusunoki H, Tsuji S, Wada Y, Fukai M, Nagai K, Itoh M, et al. Relationship between sarcopenia and the serum creatinine/cystatin C ratio in Japanese rural community-dwelling older adults. JCSM Clin Rep. 2018;3(1):1–14.

149. Savino M, Plumb L, Casula A, Evans K, Wong E, Kolhe N, et al. Acute kidney injury identification for pharmacoepidemiologic studies: Use of laboratory electronic acute kidney injury alerts versus electronic health records in Hospital Episode Statistics. Pharmacoepidemiol Drug Saf. 2021;30(12):1687–95.

150. Vlasschaert MEO, Bejaimal SAD, Hackam DG, Quinn R, Cuerden MS, Oliver MJ, et al. Validity of Administrative Database Coding for Kidney Disease: A Systematic Review. Am J Kidney Dis. 2011 Jan 1;57(1):29–43.

151. See EJ, Jayasinghe K, Glassford N, Bailey M, Johnson DW, Polkinghorne KR, et al. Long-term risk of adverse outcomes after acute kidney injury: a systematic review and meta-analysis of cohort studies using consensus definitions of exposure. Kidney Int. 2019;95(1):160–72.

152. Selby NM, Casula A, Lamming L, Stoves J, Samarasinghe Y, Lewington AJ, et al. An Organizational-Level Program of Intervention for AKI: A Pragmatic Stepped Wedge Cluster Randomized Trial. J Am Soc Nephrol. 2019 Mar 1;30(3):505–15.

153. Lachance P, Villeneuve PM, Rewa OG, Wilson FP, Selby NM, Featherstone RM, et al. Association between e-alert implementation for detection of acute kidney injury and outcomes: a systematic review. Nephrol Dial Transplant. 2017 Jan 14;gfw424.

154. Connell A, Raine R, Martin P, Barbosa EC, Morris S, Nightingale C, et al. Implementation of a Digitally Enabled Care Pathway (Part 1): Impact on Clinical Outcomes and Associated Health Care Costs. J Med Internet Res. 2019 Jul 15;21(7):e13147.

155. Liu FX, Rutherford P, Smoyer-Tomic K, Prichard S, Laplante S. A global overview of renal registries: a systematic review. BMC Nephrol. 2015 Mar 19;16:31.

156. Plumb LA, Hamilton AJ, Inward CD, Ben-Shlomo Y, Caskey FJ. Continually improving standards of care: The UK Renal Registry as a translational public health tool. Pediatr Nephrol. 2018 Mar;33(3):373–80.

157. See EJ, Bello AK, Levin A, Lunney M, Osman MA, Ye F, et al. Availability, coverage, and scope of health information systems for kidney care across world countries and regions. Nephrol Dial Transplant. 2022 Jan 1;37(1):159–67.

158. Mathes T, Buehn S, Prengel P, Pieper D. Registry-based randomized controlled trials merged the strength of randomized controlled trails and observational studies and give rise to more pragmatic trials. J Clin Epidemiol. 2018 Jan 1;93:120–7.

159. Caskey FJ, Procter S, MacNeill SJ, Wade J, Taylor J, Rooshenas L, et al. The high-volume haemodiafiltration vs high-flux haemodialysis registry trial (H4RT): a multi-centre, unblinded,

randomised, parallel-group, superiority study to compare the effectiveness and cost-effectiveness of high-volume haemodiafiltration and high-flux haemodialysis in people with kidney failure on maintenance dialysis using linkage to routine healthcare databases for outcomes. Trials. 2022 Jun 27;23(1):532.

160. Guthrie G, Guthrie B, Walker H, James MT, Selby NM, Tonelli M, et al. Developing an AKI Consensus Definition for Database Research: Findings From a Scoping Review and Expert Opinion Using a Delphi Process. Am J Kidney Dis Off J Natl Kidney Found. 2022 Apr;79(4):488-496.e1.

161. Johnson M, Hounkpatin H, Fraser S, Culliford D, Uniacke M, Roderick P. Using a linked database for epidemiology across the primary and secondary care divide: acute kidney injury. BMC Med Inform Decis Mak. 2017 Jul 11;17(1):106.

162. Lin J, Fernandez H, Shashaty MGS, Negoianu D, Testani JM, Berns JS, et al. False-positive rate of AKI using consensus creatinine–based criteria. Clin J Am Soc Nephrol. 2015;10(10):1723–31.

163. Sawhney S, Fluck N, Fraser SD, Marks A, Prescott GJ, Roderick PJ, et al. KDIGO-based acute kidney injury criteria operate differently in hospitals and the community—findings from a large population cohort. Nephrol Dial Transplant. 2016 Jun 1;31(6):922–9.

164. Syme NR, Stevens K, Stirling C, McMillan DC, Talwar D. Clinical and Analytical Impact of Moving from Jaffe to Enzymatic Serum Creatinine Methodology. J Appl Lab Med. 2020 Jul 1;5(4):631–42.

165. Boss K, Stolpe S, Müller A, Wagner B, Wichert M, Assert R, et al. Effect of serum creatinine difference between the Jaffe and the enzymatic method on kidney disease detection and staging. Clin Kidney J. 2023 Nov 1;16(11):2147–55.

166. Marrington R, MacKenzie F. Variation of eGFR reporting and CKD equations used in the United Kingdom. Ann Clin Biochem Int J Lab Med. 2023 Sep;60(5):328–38.

167. Aylward RE, van der Merwe E, Pazi S, van Niekerk M, Ensor J, Baker D, et al. Risk factors and outcomes of acute kidney injury in South African critically ill adults: a prospective cohort study. BMC Nephrol. 2019 Dec 10;20(1):460.

168. Chothia MY, Ramsunder N. In-hospital mortality of HIV-positive patients with acute kidney injury a decade after the roll-out of anti-retroviral therapy in Cape Town, South Africa. Afr J Nephrol. 2019 Oct 24;22(1):46–53.

169. Rossouw E, Chetty S. Acute kidney injury after major non-cardiac surgery: Incidence and risk factors. South Afr Med J Suid-Afr Tydskr Vir Geneeskd. 2023 Mar 2;113(3):135–40.

170. Skinner DL, Hardcastle TC, Rodseth RN, Muckart DJJ. The incidence and outcomes of acute kidney injury amongst patients admitted to a level I trauma unit. Injury. 2014 Jan;45(1):259–64.

171. Mweene MD, Richards GA, Paget G, Banda J, Dickens C. Risk factors and outcomes of sepsis-associated acute kidney injury in intensive care units in Johannesburg, South Africa. South Afr Med J Suid-Afr Tydskr Vir Geneeskd. 2022 Dec 1;112(12):919–23.

172. Conti-Ramsden FI, Nathan HL, De Greeff A, Hall DR, Seed PT, Chappell LC, et al. Pregnancy-Related Acute Kidney Injury in Preeclampsia: Risk Factors and Renal Outcomes. Hypertens Dallas Tex 1979. 2019 Nov;74(5):1144–51.

173. Dave JA, Tamuhla T, Tiffin N, Levitt NS, Ross IL, Toet W, et al. Risk factors for COVID-19 hospitalisation and death in people living with diabetes: A virtual cohort study from the Western Cape Province, South Africa. Diabetes Res Clin Pract [Internet]. 2021 Jul 1 [cited 2022 Jul 20];177. Available from: https://www.diabetesresearchclinicalpractice.com/article/S0168-8227(21)00285-0/fulltext

174. Osei-Yeboah R, Tamuhla T, Ngwenya O, Tiffin N. Accessing HIV care may lead to earlier ascertainment of comorbidities in health care clients in Khayelitsha, Cape Town. PLOS Glob Public Health. 2021 Dec 22;1(12):e0000031.

175. Tamuhla T, Dave JA, Raubenheimer P, Tiffin N. Diabetes in a TB and HIV-endemic South African population: Analysis of a virtual cohort using routine health data. PLOS ONE. 2021 May 7;16(5):e0251303.

176. Mutemaringa T, Heekes A, Smith M, Boulle A, Tiffin N. Record linkage for Routinely Collected Health Data in an African Health Information Exchange. Int J Popul Data Sci [Internet]. 2023 Mar 2 [cited 2023 Jul 17];8(1). Available from: https://ijpds.org/article/view/1771

177. Sawhney S, Bell S, Black C, Christiansen CF, Heide-Jørgensen U, Jensen SK, et al. Harmonization of epidemiology of acute kidney injury and acute kidney disease produces comparable findings across four geographic populations. Kidney Int. 2022 Jun 1;101(6):1271–81.

178. Statistics South Africa. Mid-year population estimates Statistical Release P0302. 2018; Available from: statssa.gov.za

179. Western Cape Government Cape Metro District Health Plan 2020 - 2021. westerncape.gov.za; 2020.

180. Georges TD, Marie-Patrice H, Ingrid TS, Mbua RG, Hermine FM, Gloria A. Causes and outcome of acute kidney injury amongst adults patients in two hospitals of different category in Cameroon; a 5 year retrospective comparative study. BMC Nephrol. 2022 Nov 14;23(1):364.

181. Kister TS, Schmidt M, Heuft L, Federbusch M, Haase M, Kaiser T. Laboratory diagnostic of acute kidney injury and its progression: risk of underdiagnosis in female and elderly patients [Internet]. medRxiv; 2022 [cited 2024 Feb 7]. p. 2022.05.13.22275003. Available from: https://www.medrxiv.org/content/10.1101/2022.05.13.22275003v1

182. Tambe BA, Mabapa NS, Mbhatsani HV, Mandiwana TC, Mushaphi LF, Mohlala M, et al. Household socio-economic determinants of food security in Limpopo Province of South Africa: a cross sectional survey. Agric Food Secur. 2023 Jul 17;12(1):19.

183. Swart EC, van der Merwe M, Williams J, Blaauw F, Viljoen JMM, Schenck CJ. Nutritional Status, Dietary Intake and Dietary Diversity of Landfill Waste Pickers. Nutrients. 2022 Mar 10;14(6):1172.

184. Moyo S, Ismail F, Walt MV der, Ismail N, Mkhondo N, Dlamini S, et al. Prevalence of bacteriologically confirmed pulmonary tuberculosis in South Africa, 2017–19: a multistage, cluster-based, cross-sectional survey. Lancet Infect Dis. 2022 Aug 1;22(8):1172–80.

185. Mendham AE, Goedecke JH, Micklesfield LK, Brooks NE, Faber M, Christensen DL, et al. Understanding factors associated with sarcopenic obesity in older African women from a low-income setting: a cross-sectional analysis. BMC Geriatr. 2021 Apr 14;21(1):247.

186. Berglund F, Killander J, Pompeius R. Effect of trimethoprim-sulfamethoxazole on the renal excretion of creatinine in man. J Urol. 1975 Dec;114(6):802–8.

187. Lindeman TA, Duggan JM, Sahloff EG. Evaluation of Serum Creatinine Changes With Integrase Inhibitor Use in Human Immunodeficiency Virus-1 Infected Adults. Open Forum Infect Dis. 2016 Mar 11;3(2):ofw053.

188. Peer N, Steyn K, Lombard C, Gwebushe N, Levitt N. A high burden of hypertension in the urban black population of Cape Town: the cardiovascular risk in Black South Africans (CRIBSA) study. PloS One. 2013;8(11):e78567.

189. Cartin-Ceba R, Kashiouris M, Plataki M, Kor DJ, Gajic O, Casey ET. Risk factors for development of acute kidney injury in critically ill patients: a systematic review and meta-analysis of observational studies. Crit Care Res Pract. 2012;2012:691013.

190. Boake M, Mash R. Diabetes in the Western Cape, South Africa: A secondary analysis of the diabetes cascade database 2015 – 2020. Prim Care Diabetes. 2022 Aug;16(4):555–61.

191. Johnson LF, Meyer-Rath G, Dorrington RE, Puren A, Seathlodi T, Zuma K, et al. The Effect of HIV Programs in South Africa on National HIV Incidence Trends, 2000-2019. J Acquir Immune Defic Syndr 1999. 2022 Jun 1;90(2):115–23.

192. Richard Osei-Yeboah, Olina Ngwenya, Nicki Tiffin. Kidney function in healthcare clients in Khayelitsha, South Africa: Routine laboratory testing and results reflect distinct healthcare experiences by age for healthcare clients with and without HIV. PLOS Glob Public Health. In Press.

193. Diamantidis CJ, Zepel L, Smith VA, Brookhart MA, Burks E, Bowling CB, et al. Epidemiology of Community-Acquired Acute Kidney Injury Among US Veterans. Am J Kidney Dis [Internet]. 2023 Mar 22 [cited 2023 Mar 23];0(0). Available from: https://www.ajkd.org/article/S0272-6386(23)00580-2/fulltext?dgcid=raven_jbs_aip_email

194. Silver SA, Adhikari NK, Jeyakumar N, Luo B, Harel Z, Dixon SN, et al. Association of an Acute Kidney Injury Follow-up Clinic With Patient Outcomes and Care Processes: A Cohort Study. Am J Kidney Dis [Internet]. 2022 Dec 12 [cited 2023 Feb 8];0(0). Available from: https://www.ajkd.org/article/S0272-6386(22)01052-6/fulltext

195. Davids MR, Jardine T, Marais N, Sebastian S, Jacobs JC. South African Renal Registry Annual Report 2020. 2020;

196. Fu EL, Janse RJ, de Jong Y, van der Endt VHW, Milders J, van der Willik EM, et al. Acute kidney injury and kidney replacement therapy in COVID-19: a systematic review and meta-analysis. Clin Kidney J. 2020 Aug;13(4):550–63.

197. Chu KM, Marco JL, Owolabi EO, Duvenage R, Londani M, Lombard C, et al. Trauma trends during COVID-19 alcohol prohibition at a South African regional hospital. Drug Alcohol Rev. 2022;41(1):13–9.

198. Jardim CGR, Zamani R, Akrami M. Evaluating the Impact of the COVID-19 Pandemic on Accessing HIV Services in South Africa: A Systematic Review. Int J Environ Res Public Health. 2022 Sep 21;19(19):11899.

199. Hendrikse C, Parak M, Van Hoving DJ. A descriptive analysis of the effect of the national COVID-19 lockdown on the workload and case mix of patients presenting to a district-level emergency centre in Cape Town, South Africa. South Afr Med J Suid-Afr Tydskr Vir Geneeskd. 2020 Oct 12;110(11):1113–8.

200. Navsaria PH, Nicol AJ, Parry CDH, Matzopoulos R, Maqungo S, Gaudin R. The effect of lockdown on intentional and nonintentional injury during the COVID-19 pandemic in Cape Town, South Africa: A preliminary report. S Afr Med J. 2020 Dec 14;111(2):110–3.

201. Chu KM, Marco J, Bougard H, Strauss CP, Bertels L, Victor AE, et al. Estimating the surgical backlog from the COVID-19 lockdown in South Africa: A retrospective analysis of six government hospitals. S Afr Med J. 2021 May 10;111(7):685–8.

202. Kruger EC, Banderker R, Erasmus RT, Zemlin AE. The impact of COVID-19 on routine patient care from a laboratory perspective. South Afr Med J Suid-Afr Tydskr Vir Geneeskd. 2020 Nov 5;110(12):1201–5.

203. Adeniyi AB, Laurence CE, Volmink JA, Davids MR. Prevalence of chronic kidney disease and association with cardiovascular risk factors among teachers in Cape Town, South Africa. Clin Kidney J. 2017 Jun 1;10(3):363–9.

204. Navise NH, Mokwatsi GG, Gafane-Matemane LF, Fabian J, Lammertyn L. Kidney dysfunction: prevalence and associated risk factors in a community-based study from the North West Province of South Africa. BMC Nephrol. 2023 Jan 30;24(1):23.

205. Nel J, Dlamini S, Meintjes G, Burton R, Black JM, Davies NECG, et al. Southern African HIV Clinicians Society guidelines for antiretroviral therapy in adults: 2020 update. South Afr J HIV Med. 2020;21(1):1115.

206. Matsha TE, Yako YY, Rensburg MA, Hassan MS, Kengne AP, Erasmus RT. Chronic kidney diseases in mixed ancestry south African populations: prevalence, determinants and concordance between kidney function estimators. BMC Nephrol. 2013 Apr 2;14:75.

207. Peer N, George J, Lombard C, Steyn K, Levitt N, Kengne AP. Prevalence, concordance and associations of chronic kidney disease by five estimators in South Africa. BMC Nephrol. 2020 Aug 27;21(1):372.

208. Matsushita K, Mahmoodi BK, Woodward M, Emberson JR, Jafar TH, Jee SH, et al. Comparison of Risk Prediction Using the CKD-EPI Equation and the MDRD Study Equation for Estimated Glomerular Filtration Rate. JAMA. 2012 May 9;307(18):1941–51.

209. Moosa MR. The state of kidney transplantation in South Africa. South Afr Med J Suid-Afr Tydskr Vir Geneeskd. 2019 Mar 29;109(4):235–40.

210. Moosa MR, Kidd M. The dangers of rationing dialysis treatment: The dilemma facing a developing country. Kidney Int. 2006 Sep 2;70(6):1107–14.

211. Carrero JJ, Fu EL, Vestergaard SV, Jensen SK, Gasparini A, Mahalingasivam V, et al. Defining measures of kidney function in observational studies using routine healthcare data:

methodological and reporting considerations. Kidney Int [Internet]. 2022 Oct 21 [cited 2022 Oct 28];0(0). Available from: https://www.kidney-international.org/article/S0085-2538(22)00842-0/fulltext?dgcid=raven_jbs_aip_email

212. McDonald HI, Shaw C, Thomas SL, Mansfield KE, Tomlinson LA, Nitsch D. Methodological challenges when carrying out research on CKD and AKI using routine electronic health records. Kidney Int. 2016 Nov 1;90(5):943–9.

213. Sykes L, Asar O, Ritchie J, Raman M, Vassallo D, Alderson HV, et al. The influence of multiple episodes of acute kidney injury on survival and progression to end stage kidney disease in patients with chronic kidney disease. PLOS ONE. 2019 Jul 18;14(7):e0219828.

214. Sawhney S, Mitchell M, Marks A, Fluck N, Black C. Long-term prognosis after acute kidney injury (AKI): what is the role of baseline kidney function and recovery? A systematic review. BMJ Open. 2015 Jan 1;5(1):e006497.

215. Hsu RK, Hsu CY, McCulloch CE, Yang J, Anderson AH, Chen J, et al. Research-based versus clinical serum creatinine measurements and the association of acute kidney injury with subsequent kidney function: findings from the Chronic Renal Insufficiency Cohort study. Clin Kidney J. 2020 Feb;13(1):55–62.

216. Hamroun A, Frimat L, Laville M, Metzger M, Combe C, Fouque D, et al. New insights into acute-on-chronic kidney disease in nephrology patients: the CKD-REIN study. Nephrol Dial Transplant. 2022 Sep 1;37(9):1700–9.

217. Hatakeyama Y, Horino T, Kataoka H, Matsumoto T, Ode K, Shimamura Y, et al. Incidence of acute kidney injury among patients with chronic kidney disease: a single-center retrospective database analysis. Clin Exp Nephrol. 2017;21(1):43–8.

218. Hsu C yuan, Chertow GM, McCulloch CE, Fan D, Ordoñez JD, Go AS. Nonrecovery of kidney function and death after acute on chronic renal failure. Clin J Am Soc Nephrol CJASN. 2009 May;4(5):891–8.

219. Stengel B, Combe C, Jacquelinet C, Briançon S, Fouque D, Laville M, et al. The French Chronic Kidney Disease-Renal Epidemiology and Information Network (CKD-REIN) cohort study. Nephrol Dial Transplant. 2014 Aug 1;29(8):1500–7.

220. Ren L, Li F, Di Z, Xiong Y, Zhang S, Ma Q, et al. Estradiol Ameliorates Acute Kidney Ischemia-Reperfusion Injury by Inhibiting the TGF-βRI-SMAD Pathway. Front Immunol. 2022;13:822604.

221. D'hoore E, Neirynck N, Schepers E, Vanholder R, Verbeke F, Van Thielen M, et al. Chronic kidney disease progression is mainly associated with non-recovery of acute kidney injury. J Nephrol. 2015;

222. Hsu CY, Chertow GM, McCulloch CE, Fan D, Ordoñez JD, Go AS. Nonrecovery of kidney function and death after acute on chronic renal failure. Clin J Am Soc Nephrol. 2009;4(5):891–8.

223. Tannor EK, Chika OU, Okpechi IG. The Impact of Low Socioeconomic Status on Progression of Chronic Kidney Disease in Low- and Lower Middle-Income Countries. Semin Nephrol [Internet]. 2022 Sep 1 [cited 2023 Jun 13];42(5). Available from: https://www.seminarsinnephrology.org/article/S0270-9295(23)00048-7/fulltext?dgcid=raven_jbs_etoc_email

224. Alaba O, Chola L. The social determinants of multimorbidity in South Africa. Int J Equity Health. 2013 Aug 20;12:63.

225. Larsen T, See EJ, Holmes NE, Bellomo R. Estimating baseline creatinine to detect acute kidney injury in patients with chronic kidney disease. Nephrology. 2023;28(8):434–45.

226. Xu X, Wang L, Xiong M, Nie S, Liu Z, Feng Y, et al. New Criterion to Evaluate Acute-on-Chronic Kidney Injury Based on the Creatinine Reference Change. Am J Nephrol. 2020;51(6):453–62.

227. Warnock DG, Neyra JA, Macedo E, Miles AD, Mehta RL, Wanner C. Comparison of Static and Dynamic Baseline Creatinine Surrogates for Defining Acute Kidney Injury. Nephron. 2021;145(6):664–74.

228. Kuligowski J, Pérez-Guaita D, Quintás G. Application of Discriminant Analysis and Cross-Validation on Proteomics Data. In: Jung K, editor. Statistical Analysis in Proteomics [Internet]. New York, NY: Springer; 2016 [cited 2023 Feb 22]. p. 175–84. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-4939-3106-4_11

229. Tilford CA, Siemers NO. Gene set enrichment analysis. Methods Mol Biol Clifton NJ. 2009;563:99–121.

230. Li L, He M, Tang X, Huang J, Li J, Hong X, et al. Proteomic landscape of the extracellular matrix in the fibrotic kidney. Kidney Int [Internet]. 2023 Feb 17 [cited 2023 Feb 19];0(0). Available from: https://www.kidney-international.org/article/S0085-2538(23)00082-0/fulltext?dgcid=raven_jbs_aip_email

231. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018 Jan 4;46(D1):D649–55.

232. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022 Jan 7;50(D1):D687–92.

233. Ringnér M. What is principal component analysis? Nat Biotechnol. 2008 Mar;26(3):303–4.

234. Principal Component Analysis [Internet]. New York: Springer-Verlag; 2002 [cited 2023 Jan 26]. (Springer Series in Statistics). Available from: http://link.springer.com/10.1007/b98835

235. Cattell RB. The Scree Test For The Number Of Factors. Multivar Behav Res. 1966 Apr 1;1(2):245–76.

236. Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 2000 Sep 15;28(18):3442–4.

237. Gansevoort RT, Anders HJ, Cozzolino M, Fliser D, Fouque D, Ortiz A, et al. What should European nephrology do with the new CKD-EPI equation? Nephrol Dial Transplant. 2022 Sep 7;gfac254.

238. Diez-Roux AV. Multilevel analysis in public health research. Annu Rev Public Health. 2000;21:171–92.

239. Rönkkö M, Aalto E, Tenhunen H, Aguirre-Urreta MI. Eight Simple Guidelines for Improved Understanding of Transformations and Nonlinear Effects. Organ Res Methods. 2022 Jan 1;25(1):48–87.

240. Wooldridge JM. Econometric analysis of cross section and panel data. 2nd ed. Cambridge, Mass: MIT Press; 2010. 1064 p.

241. Pek J, Wong O, Wong AC. Data Transformations for Inference with Linear Regression: Clarifications and Recommendations. [cited 2023 Feb 21]; Available from: https://scholarworks.umass.edu/pare/vol22/iss1/9/

242. Mansournia MA, Nazemipour M, Naimi AI, Collins GS, Campbell MJ. Reflection on modern methods: demystifying robust standard errors for epidemiologists. Int J Epidemiol. 2021 Mar 3;50(1):346–51.

243. Littell RC, Pendergast J, Natarajan R. Modelling covariance structure in the analysis of repeated measures data. Stat Med. 2000 Jul 15;19(13):1793–819.

244. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiol Camb Mass. 1999 Jan;10(1):37–48.

245. PEARL J. Causal diagrams for empirical research. Biometrika. 1995 Dec 1;82(4):669–88.

246. Collider Bias - PubMed [Internet]. [cited 2022 Dec 21]. Available from: https://pubmed-ncbi-nlm-nih-gov.bris.idm.oclc.org/35285854/

247. Venkat-Raman G, Tomson CRV, Gao Y, Cornet R, Stengel B, Gronhagen-Riska C, et al. New primary renal diagnosis codes for the ERA-EDTA. Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc. 2012 Dec;27(12):4414–9.

248. Major RW, Cheng MRI, Grant RA, Shantikumar S, Xu G, Oozeerally I, et al. Cardiovascular disease risk factors in chronic kidney disease: A systematic review and meta-analysis. PLoS ONE. 2018;

249. Remuzzi G, Perico N, Macia M, Ruggenenti P. The role of renin-angiotensin-aldosterone system in the progression of chronic kidney disease. Kidney Int. 2005 Dec 1;68:S57–65.

250. Randomised placebo-controlled trial of effect of ramipril on decline in glomerular filtration rate and risk of terminal renal failure in proteinuric, non-diabetic nephropathy. The Lancet. 1997 Jun 28;349(9069):1857–63.

251. Bakris GL, Hart P, Ritz E. Beta blockers in the management of chronic kidney disease. Kidney Int. 2006 Dec;70(11):1905–13.

252. Textor J, Hardt J, Knüppel S. DAGitty: A Graphical Tool for Analyzing Causal Diagrams. Epidemiology. 2011 Sep;22(5):745.

253. Schluchter MD, Greene T, Beck GJ. Analysis of change in the presence of informative censoring: Application to a longitudinal clinical trial of progressive renal disease. Stat Med. 2001;

254. Schluchter MD. Methods for the analysis of informatively censored longitudinal data. Stat Med. 1992 Nov;11(14–15):1861–70.

255. Chesnaye NC, Tripepi G, Dekker FW, Zoccali C, Zwinderman AH, Jager KJ. An introduction to joint models-applications in nephrology. Clin Kidney J. 2020 Apr;13(2):143–9.

256. Agarwal R, Bunaye Z, Bekele DM, Light RP. Competing Risk Factor Analysis of End-Stage Renal Disease and Mortality in Chronic Kidney Disease. Am J Nephrol. 2008;28(4):569–75.

257. Ravani P, Fiocco M, Liu P, Quinn RR, Hemmelgarn B, James M, et al. Influence of Mortality on Estimating the Risk of Kidney Failure in People with Stage 4 CKD. J Am Soc Nephrol. 2019 Nov;30(11):2219–27.

258. Teixeira L, Rodrigues A, Carvalho MJ, Cabrita A, Mendonça D. Modelling competing risks in nephrology research: an example in peritoneal dialysis. BMC Nephrol. 2013 Dec;14(1):110.

259. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J Am Stat Assoc. 1999 Jun;94(446):496–509.

260. Liu L, Ma JZ, O'Quigley J. Joint analysis of multi-level repeated measures data and survival: An application to the end stage renal disease (ESRD) data. Stat Med. 2008;

261. Yang W, Xie D, Pan Q, Feldman HI, Guo W. Joint Modeling of Repeated Measures and Competing Failure Events in a Study of Chronic Kidney Disease. Stat Biosci. 2017 Dec;9(2):504–24.

262. Crowther MJ. merlin—A unified modeling framework for data analysis and methods development in Stata. Stata J. 2020 Dec 1;20(4):763–84.

263. Crowther MJ, Abrams KR, Lambert PC. Joint Modeling of Longitudinal and Survival Data. Stata J Promot Commun Stat Stata. 2013 Mar;13(1):165–84.

264. van Walraven C, McCudden C, Austin PC. Imputing missing laboratory results may return erroneous values because they are not missing at random. J Clin Epidemiol. 2023 Feb 1;154:65–74.

265. Weaver RG, James MT, Ravani P, Weaver CGW, Lamb EJ, Tonelli M, et al. Estimating Urine Albumin-to-Creatinine Ratio from Protein-to-Creatinine Ratio: Development of Equations using Same-Day Measurements. J Am Soc Nephrol JASN. 2020 Mar;31(3):591–601.

266. RUBIN DB. Inference and missing data. Biometrika. 1976 Dec 1;63(3):581–92.

267. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. Int J Epidemiol. 2019 Aug 1;48(4):1294–304.

268. Blazek K, van Zwieten A, Saglimbene V, Teixeira-Pinto A. A practical guide to multiple imputation of missing data in nephrology. Kidney Int. 2021 Jan 1;99(1):68–74.

269. Rubin DB, Schenker N. Multiple Imputation for Interval Estimation From Simple Random Samples With Ignorable Nonresponse. J Am Stat Assoc. 1986;81(394):366–74.

270. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377–99.

271. Lee KJ, Carlin JB. Multiple imputation in the presence of non-normal data. Stat Med. 2017 Feb 20;36(4):606–17.

272. Tompsett DM, Leacy F, Moreno-Betancur M, Heron J, White IR. On the use of the not-at-random fully conditional specification (NARFCS) procedure in practice. Stat Med. 2018 Jul 10;37(15):2338–53.

273. Quartagno M, Grund S, Carpenter J. jomo: A Flexible Package for Two-level Joint Modelling Multiple Imputation. R J. 2019;11(2):205–28.

274. Enders CK, Mistler SA, Keller BT. Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. Psychol Methods. 2016 Jun;21(2):222–40.

275. R B, S L. Adjusting for multiple testing--when and how? J Clin Epidemiol [Internet]. 2001 Apr [cited 2022 Jul 29];54(4). Available from: https://pubmed.ncbi.nlm.nih.gov/11297884/

276. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilità [Internet]. Seeber; 1936. Available from: https://books.google.co.za/books?id=3CY-HQAACAAJ

277. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.

278. Roger Newson. Multiple-test procedures and smile plots. Stata J. 2003;3(2):109–32.

279. Lv L, Wang J, Gao B, Wu L, Wang F, Cui Z, et al. Serum uromodulin and progression of kidney disease in patients with chronic kidney disease. J Transl Med. 2018 Dec;16(1):316.

280. Scherberich JE, Gruber R, Nockher WA, Christensen EI, Schmitt H, Herbst V, et al. Serum uromodulin—a marker of kidney function and renal parenchymal integrity. Nephrol Dial Transplant. 2018 Feb 1;33(2):284–95.

281. Scholze A, Krueger K, Diedrich M, Räth C, Torges A, Jankowski V, et al. Superoxide dismutase type 1 in monocytes of chronic kidney disease patients. Amino Acids. 2011 Jul;41(2):427–38.

282. Hale AJ, ter Steege E, den Hertog J. Recent advances in understanding the role of protein-tyrosine phosphatases in development and disease. Dev Biol. 2017 Aug 15;428(2):283–92.

283. Aricescu AR, McKinnell IW, Halfter W, Stoker AW. Heparan sulfate proteoglycans are ligands for receptor protein tyrosine phosphatase sigma. Mol Cell Biol. 2002 Mar;22(6):1881–92.

284. Termini CM, Pang A, Batton DM, Chute JP. Proteoglycans regulate protein tyrosine phosphatase receptor σ organization on hematopoietic stem/progenitor cells. Exp Hematol. 2021 Apr 1;96:44–51.

285. Yan H, Grossman A, Wang H, D'Eustachio P, Mossie K, Musacchio JM, et al. A novel receptor tyrosine phosphatase-sigma that is highly expressed in the nervous system. J Biol Chem. 1993 Nov 25;268(33):24880–6.

286. Muise AM, Walters T, Wine E, Griffiths AM, Turner D, Duerr RH, et al. Protein-tyrosine phosphatase sigma is associated with ulcerative colitis. Curr Biol CB. 2007 Jul 17;17(14):1212–8.

287. Doody KM, Stanford SM, Sacchetti C, Svensson MND, Coles CH, Mitakidis N, et al. Targeting phosphatase-dependent proteoglycan switch for rheumatoid arthritis therapy. Sci Transl Med. 2015 May 20;7(288):288ra76.

288. Uetani N, Bertozzi K, Chagnon MJ, Hendriks W, Tremblay ML, Bouchard M. Maturation of ureter-bladder connection in mice is controlled by LAR family receptor protein tyrosine phosphatases. J Clin Invest. 2009 Apr 1;119(4):924–35.

289. Berillo O, Huo KG, Richer C, Fraulob-Aquino JC, Briet M, Lipman ML, et al. Distinct transcriptomic profile of small arteries of hypertensive patients with chronic kidney disease identified miR-338-3p targeting GPX3 and PTPRS. J Hypertens. 2022 Jul;40(7):1394–405.

290. Thompson KM, Uetani N, Manitt C, Elchebly M, Tremblay ML, Kennedy TE. Receptor protein tyrosine phosphatase sigma inhibits axonal regeneration and the rate of axon extension. Mol Cell Neurosci. 2003 Aug;23(4):681–92.

291. Gardner RT, Habecker BA. Infarct-derived chondroitin sulfate proteoglycans prevent sympathetic reinnervation after cardiac ischemia-reperfusion injury. J Neurosci Off J Soc Neurosci. 2013 Apr 24;33(17):7175–83.

292. Grassi G, Biffi A, Seravalle G, Bertoli S, Airoldi F, Corrao G, et al. Sympathetic nerve traffic overactivity in chronic kidney disease: a systematic review and meta-analysis. J Hypertens. 2021 Mar 1;39(3):408–16.

293. Jang HS, Kim J, Padanilam BJ. Renal sympathetic nerve activation via α_2-adrenergic receptors in chronic kidney disease progression. Kidney Res Clin Pract. 2019 Mar 4;38(1):6–14.

294. Shao X, Taha IN, Clauser KR, Gao Y (Tom), Naba A. MatrisomeDB: the ECM-protein knowledge database. Nucleic Acids Res. 2020 Jan 8;48(D1):D1136–44.

295. Schmidt IM, Colona MR, Kestenbaum BR, Alexopoulos LG, Palsson R, Srivastava A, et al. Cadherin-11, Sparc-related modular calcium binding protein-2, and Pigment epithelium-derived factor are promising non-invasive biomarkers of kidney fibrosis. Kidney Int. 2021 Sep;100(3):672–83.

296. Dendooven A, van Oostrom O, van der Giezen DM, Leeuwis JW, Snijckers C, Joles JA, et al. Loss of endogenous bone morphogenetic protein-6 aggravates renal fibrosis. Am J Pathol. 2011 Mar;178(3):1069–79.

297. Cai G, Zhang X, Hong Q, Shao F, Shang X, Fu B, et al. Tissue inhibitor of metalloproteinase-1 exacerbated renal interstitial fibrosis through enhancing inflammation. Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc. 2008 Jun;23(6):1861–75.

298. Ruiz-Ortega M, Lamas S, Ortiz A. Antifibrotic Agents for the Management of CKD: A Review. Am J Kidney Dis. 2022 Aug 1;80(2):251–63.

299. Ito S, Ozaki T, Morozumi M, Imagama S, Kadomatsu K, Sakamoto K. Enoxaparin promotes functional recovery after spinal cord injury by antagonizing PTPRσ. Exp Neurol. 2021 Jun 1;340:113679.

300. Kappert K, Peters KG, Böhmer FD, Östman A. Tyrosine phosphatases in vessel wall signaling. Cardiovasc Res. 2005 Feb 15;65(3):587–98.

301. Vercauteren M, Remy E, Devaux C, Dautreaux B, Henry JP, Bauer F, et al. Improvement of Peripheral Endothelial Dysfunction by Protein Tyrosine Phosphatase Inhibitors in Heart Failure. Circulation. 2006 Dec 5;114(23):2498–507.

302. Chen JX, Tuo Q, Liao DF, Zeng H. Inhibition of Protein Tyrosine Phosphatase Improves Angiogenesis via Enhancing Ang-1/Tie-2 Signaling in Diabetes. J Diabetes Res. 2012 Feb 12;2012:e836759.

303. Dubin RF, Deo R, Ren Y, Wang J, Zheng Z, Shou H, et al. Proteomics of CKD progression in the chronic renal insufficiency cohort. Nat Commun. 2023 Oct 10;14(1):6340.

304. Norris K, Norris F, Kono DH, Vestergaard H, Pedersen O, Theofilopoulos AN, et al. Expression of protein-tyrosine phosphatases in the major insulin target tissues. FEBS Lett. 1997;415(3):243–8.

305. Långberg EC, Gu HF, Nordman S, Efendic S, Östenson CG. Genetic variation in receptor protein tyrosine phosphatase σ is associated with type 2 diabetes in Swedish Caucasians. Eur J Endocrinol. 2007 Oct 1;157(4):459–64.

306. Zhu Y, Chen L, Song B, Cui Z, Chen G, Yu Z, et al. Insulin-like Growth Factor-2 (IGF-2) in Fibrosis. Biomolecules. 2022 Oct 25;12(11):1557.

307. Liso A, Venuto S, Coda ARD, Giallongo C, Palumbo GA, Tibullo D. IGFBP-6: At the Crossroads of Immunity, Tissue Repair and Fibrosis. Int J Mol Sci. 2022 Apr 14;23(8):4358.

308. el Nahas AM. The role of growth hormone and insulin-like growth factor-I in experimental renal growth and scarring. Am J Kidney Dis Off J Natl Kidney Found. 1991 Jun;17(6):677–9.

309. Kelley KM, Oh Y, Gargosky SE, Gucev Z, Matsumoto T, Hwa V, et al. Insulin-like growth factor-binding proteins (IGFBPs) and their regulatory dynamics. Int J Biochem Cell Biol. 1996 Jun 1;28(6):619–37.

310. Grellier P, Sabbah M, Fouqueray B, Woodruff K, Yee D, Abboud HE, et al. Characterization of insulin-like growth factor binding proteins and regulation of IGFBP3 in human mesangial cells. Kidney Int. 1996 Apr;49(4):1071–8.

311. Johnson DW, Saunders HJ, Brew BK, Ganesan A, Baxter RC, Poronnik P, et al. Human renal fibroblasts modulate proximal tubule cell growth and transport via the IGF-I axis. Kidney Int. 1997 Dec;52(6):1486–96.

312. Dittmann K, Wallaschofski H, Rettig R, Stracke S, Endlich K, Völzke H, et al. Association between serum insulin-like growth factor I or IGF-binding protein 3 and estimated glomerular filtration rate: results of a population-based sample. BMC Nephrol. 2012 Dec 13;13(1):169.

313. Powell DR, Liu F, Baker BK, Hintz RL, Durham SK, Brewer ED, et al. Insulin-like growth factor-binding protein-6 levels are elevated in serum of children with chronic renal failure: a report of the Southwest Pediatric Nephrology Study Group. J Clin Endocrinol Metab. 1997 Sep;82(9):2978–84.

314. Brittain AL, Kopchick JJ. A review of renal GH/IGF1 family gene expression in chronic kidney diseases. Growth Horm IGF Res. 2019 Oct;48–49:1–4.

315. Fukuda I, Hizuka N, Okubo Y, Takano K, Asakawa-Yasumoto K, Shizume K, et al. Changes in serum insulin-like growth factor binding protein-2, -3, and -6 levels in patients with chronic renal failure following renal transplantation. Growth Horm IGF Res. 1998 Dec 1;8(6):481–6.

316. Andersson P, Gustafsson T, Arnqvist HJ. Insulin-like growth factor binding proteins-2 to -6 are expressed by human vascular smooth muscle cells. J Endocrinol. 1999 Nov;163(2):281–8.

317. Price GJ, Berka JL, Edmondson SR, Werther GA, Bach LA. Localization of mRNAs for insulin-like growth factor binding proteins 1 to 6 in rat kidney. Kidney Int. 1995 Aug;48(2):402–11.

318. Fu H, Tian Y, Zhou L, Zhou D, Tan RJ, Stolz DB, et al. Tenascin-C Is a Major Component of the Fibrogenic Niche in Kidney Fibrosis. J Am Soc Nephrol JASN. 2017 Mar;28(3):785–801.

319. Lennon R, Byron A, Humphries JD, Randles MJ, Carisey A, Murphy S, et al. Global analysis reveals the complexity of the human glomerular extracellular matrix. J Am Soc Nephrol JASN. 2014 May;25(5):939–51.

320. Hale L, Welsh G, Perks C, Hurcombe J, Moore S, Hers I, et al. Insulin-like growth factor-II is produced by, signals to and is an important survival factor for the mature podocyte in man and mouse. J Pathol. 2013;230(1):95–106.

321. Januzzi JL Jr, Butler J, Sattar N, Xu J, Shaw W, Rosenthal N, et al. Insulin-Like Growth Factor Binding Protein 7 Predicts Renal and Cardiovascular Outcomes in the Canagliflozin Cardiovascular Assessment Study. Diabetes Care. 2020 Nov 6;44(1):210–6.

322. Noris M, Remuzzi G. Overview of Complement Activation and Regulation. Semin Nephrol. 2013 Nov 1;33(6):479–92.

323. Thurman JM. Complement in Kidney Disease: Core Curriculum 2015. Am J Kidney Dis. 2015 Jan 1;65(1):156–68.

324. Zhang XL, Ali MAM. Ficolins: structure, function and associated diseases. Adv Exp Med Biol. 2008;632:105–15.

325. Tanha N, Pilely K, Faurschou M, Garred P, Jacobsen S. Plasma ficolin levels and risk of nephritis in Danish patients with systemic lupus erythematosus. Clin Rheumatol. 2017 Feb;36(2):335–41.

326. Nielsen TL, Pilely K, Lund KP, Warming PE, Plesner LL, Iversen KK, et al. Hemodialysis leads to plasma depletion of lectin complement pathway initiator molecule ficolin-2. Hemodial Int Int Symp Home Hemodial. 2021 Oct;25(4):479–88.

327. Dabrowska-Zamojcin E, Czerewaty M, Malinowski D, Tarnowski M, Słuczanowska-Głabowska S, Domanski L, et al. Ficolin-2 Gene rs7851696 Polymorphism is Associated with Delayed Graft Function and Acute Rejection in Kidney Allograft Recipients. Arch Immunol Ther Exp (Warsz). 2018 Feb;66(1):65–72.

328. Matsushita M, Fujita T. The lectin pathway. Res Immunol. 1996 Feb;147(2):115–8.

329. Békássy ZD, Kristoffersson AC, Rebetz J, Tati R, Olin AI, Karpman D. Aliskiren inhibits renin-mediated complement activation. Kidney Int. 2018 Oct 1;94(4):689–700.

330. Zhou X, Fukuda N, Matsuda H, Endo M, Wang X, Saito K, et al. Complement 3 activates the renal renin-angiotensin system by induction of epithelial-to-mesenchymal transition of the nephrotubulus in mice. Am J Physiol Renal Physiol. 2013 Oct 1;305(7):F957-967.

331. Klahr S, Ishidoya S, Morrissey J. Role of angiotensin II in the tubulointerstitial fibrosis of obstructive nephropathy. Am J Kidney Dis Off J Natl Kidney Found. 1995 Jul;26(1):141–6.

332. Boor P, Konieczny A, Villa L, Schult AL, Bücher E, Rong S, et al. Complement C5 mediates experimental tubulointerstitial fibrosis. J Am Soc Nephrol JASN. 2007 May;18(5):1508–15.

333. Legendre CM, Licht C, Muus P, Greenbaum LA, Babu S, Bedrosian C, et al. Terminal complement inhibitor eculizumab in atypical hemolytic-uremic syndrome. N Engl J Med. 2013 Jun 6;368(23):2169–81.

334. Paréj K, Dobó J, Závodszky P, Gál P. The control of the complement lectin pathway activation revisited: Both C1-inhibitor and antithrombin are likely physiological inhibitors, while α2-macroglobulin is not. Mol Immunol. 2013 Jul 1;54(3):415–22.

335. Giraudi PJ, Salvoza N, Bonazza D, Saitta C, Lombardo D, Casagranda B, et al. Ficolin-2 Plasma Level Assesses Liver Fibrosis in Non-Alcoholic Fatty Liver Disease. Int J Mol Sci. 2022 Mar 4;23(5):2813.

336. Gaya da Costa M, Poppelaars F, Berger SP, Daha MR, Seelen MA. The lectin pathway in renal disease: old concept and new insights. Nephrol Dial Transplant. 2018 Dec 1;33(12):2073–9.

337. Roos A, Rastaldi MP, Calvaresi N, Oortwijn BD, Schlagwein N, van Gijlswijk-Janssen DJ, et al. Glomerular activation of the lectin pathway of complement in IgA nephropathy is associated with more severe renal disease. J Am Soc Nephrol JASN. 2006 Jun;17(6):1724–34.

338. Keshawarz A, Hwang SJ, Lee GY, Yu Z, Yao C, Köttgen A, et al. Cardiovascular disease protein biomarkers are associated with kidney function: The Framingham Heart Study. PloS One. 2022;17(5):e0268293.

339. Ying WZ, Li X, Rangarajan S, Feng W, Curtis LM, Sanders PW. Immunoglobulin light chains generate proinflammatory and profibrotic kidney injury. J Clin Invest. 129(7):2792–806.

340. Caza TN, Hassen SI, Kuperman M, Sharma SG, Dvanajscak Z, Arthur J, et al. Neural cell adhesion molecule 1 is a novel autoantigen in membranous lupus nephritis. Kidney Int. 2021 Jul 1;100(1):171–81.

341. Marković-Lipkovski J, Životić M, Müller CA, Tampe B, Ćirović S, Vještica J, et al. Variable Expression of Neural Cell Adhesion Molecule Isoforms in Renal Tissue: Possible Role in Incipient Renal Fibrosis. PloS One. 2015;10(9):e0137028.

342. Marković-Lipkovski J, Müller CA, Klein G, Flad T, Klatt T, Blaschke S, et al. Neural cell adhesion molecule expression on renal interstitial cells. Nephrol Dial Transplant. 2007 Jun 1;22(6):1558–66.

343. Özkan G, Güzel S, Atar RV, Fidan Ç, Kara SP, Ulusoy Ş. Elevated serum levels of procollagen C-proteinase enhancer-1 in patients with chronic kidney disease is associated with a declining glomerular filtration rate. Nephrol Carlton Vic. 2019 Sep;24(9):938–42.

344. El-Koraie AF, Baddour NM, Adam AG, El Kashef EH, El Nahas AM. Role of stem cell factor and mast cells in the progression of chronic glomerulonephritides. Kidney Int. 2001 Jul;60(1):167–72.

345. Eldewi DM, Alhabibi AM, El Sayed HME, Mahmoud SAK, El Sadek SM, Gouda RM, et al. Expression levels of complement regulatory proteins (CD35, CD55 and CD59) on peripheral blood cells of patients with chronic kidney disease. Int J Gen Med. 2019 Sep 16;12:343–51.

346. Dubin R, Cushman M, Folsom AR, Fried LF, Palmas W, Peralta CA, et al. Kidney function and multiple hemostatic markers: cross sectional associations in the multi-ethnic study of atherosclerosis. BMC Nephrol. 2011 Jan 26;12(1):3.

347. Huang MJ, Wei R bao, Wang Y, Su T yu, Di P, Li Q ping, et al. Blood coagulation system in patients with chronic kidney disease: a prospective observational study. BMJ Open. 2017 Jun 1;7(5):e014294.

348. Peralta CA, Shlipak MG, Judd S, Cushman M, McClellan W, Zakai NA, et al. Detection of Chronic Kidney Disease With Creatinine, Cystatin C, and Urine Albumin-to-Creatinine Ratio and Association With Progression to End-Stage Renal Disease and Mortality. JAMA. 2011 Apr 20;305(15):1545–52.

349. Shlipak MG, Sarnak MJ, Katz R, Fried LF, Seliger SL, Newman AB, et al. Cystatin C and the Risk of Death and Cardiovascular Events among Elderly Persons. N Engl J Med. 2005 May 19;352(20):2049–60.

350. Lafrance JP, Miller DR. Defining Acute Kidney Injury in Database Studies: The Effects of Varying the Baseline Kidney Function Assessment Period and Considering CKD Status. Am J Kidney Dis. 2010 Oct;56(4):651–60.

351. De Rosa S, Samoni S, Ronco C. Creatinine-based definitions: From baseline creatinine to serum creatinine adjustment in intensive care. Crit Care. 2016;20(1):1–5.

352. Sawhney S, Mitchell M, Marks A, Fluck N, Black C. Long-term prognosis after acute kidney injury (AKI): what is the role of baseline kidney function and recovery? A systematic review. BMJ Open. 2015 Jan 1;5(1):e006497.

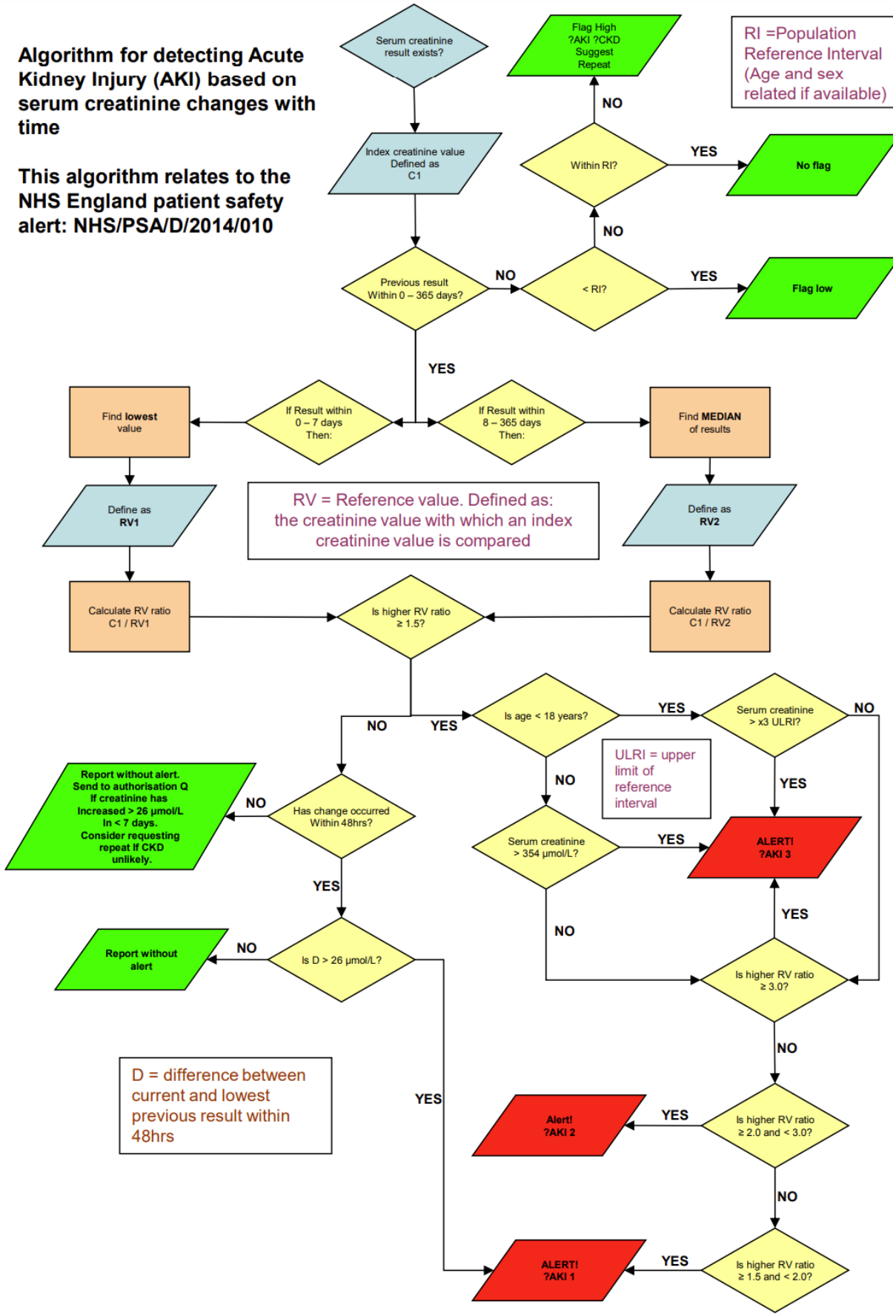
353. Mohamadlou H, Lynn-Palevsky A, Barton C, Chettipally U, Shieh L, Calvert J, et al. Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data. Can J Kidney Health Dis. 2018;5.

354. Churpek MM, Carey KA, Edelson DP, Singh T, Astor BC, Gilbert ER, et al. Internal and External Validation of a Machine Learning Risk Score for Acute Kidney Injury. JAMA Netw Open. 2020 Aug 11;3(8):e2012892.

355. Coşkun A, Sandberg S, Unsal I, Cavusoglu C, Serteser M, Kilercik M, et al. Personalized Reference Intervals in Laboratory Medicine: A New Model Based on Within-Subject Biological Variation. Clin Chem. 2021 Jan 30;67(2):374–84.

APPENDIX LIST

1 APPENDIX 1: NATIONAL HEALTH SERVICES ENGLAND ACUTE KIDNEY INJURY DETECTION ALGORITHM



Short name	UniProt ID	Long name	Panel
ACE2	Q9BYF1	Angiotensin-converting enzyme 2	CVII
ADAM-TS13	Q76LX8	A disintegrin and metalloproteinase with thrombospondin motifs 13	CVII
ADM	P35318	ADM	CVII
AGRP	O00253	Agouti-related protein	CVII
AMBP	P02760	Protein AMBP	CVII
ANG	P03950	Angiogenin	CM
ANG-1	Q15389	Angiopoietin-1	CVII
ANGPTL3	Q9Y5C1	Angiopoietin-related protein 3	CM
AOC3	Q16853	Membrane primary amine oxidase	CM
APOM	O95445	Apolipoprotein M	CM
BMP-6	P22004	Bone morphogenetic protein 6	CVII
BNP	P16860	Natriuretic peptides B	CVII
C1QTNF1	Q9BXJ1	Complement C1q tumour necrosis factor-related protein 1	CM
C2	P06681	Complement C2	CM
CA1	P00915	Carbonic anhydrase 1	CM
CA3	P07451	Carbonic anhydrase 3	CM
CA4	P22748	Carbonic anhydrase 4	CM
CA5A	P35218	Carbonic anhydrase 5A, mitochondrial	CVII
CCL14	Q16627	C-C motif chemokine 14	CM
CCL17	Q92583	C-C motif chemokine 17	CVII
CCL18	P55774	C-C motif chemokine 18	CM
CCL3	P10147	C-C motif chemokine 3	CVII
CCL5	P13501	C-C motif chemokine 5	CM
CD4	P01730	T-cell surface glycoprotein CD4	CVII

CD40-L	P29965	CD40 ligand	CVII
CD46	P15529	Membrane cofactor protein	CM
CD59	P13987	CD59 glycoprotein	CM
CD84	Q9UIB8	SLAM family member 5	CVII
CDH1	P12830	Cadherin-1	CM
CEACAM8	P31997	Carcinoembryonic antigen related cell adhesion molecule 8	CVII
CES1*	P23141	Liver carboxylesterase 1	CM
CFHR5	Q9BXR6	Complement factor H-related protein 5	CM
CHL1	O00533	Neural cell adhesion molecule L1-like protein	CM
CNDP1	Q96KN2	Beta-Ala-His dipeptidase	CM
COL18A1	P39060	Collagen alpha-1(XVIII) chain	CM
COMP	P49747	Cartilage oligomeric matrix protein	CM
CR2	P20023	Complement receptor type 2	CM
CRTAC1	Q9NQ79	Cartilage acidic protein 1	CM
CST3	P01034	Cystatin-C	CM
CTRC	Q99895	Chymotrypsin C	CVII
CTSL1	P07711	Cathepsin L1	CVII
CXCL1	P09341	C-X-C motif chemokine 1	CVII
DCN	P07585	Decorin	CVII
DECR1	Q16698	2,4-dienoyl-CoA reductase, mitochondrial	CVII
DEFA1*	P59665	Neutrophil defensin 1	CM
Dkk-1	O94907	Dickkopf-related protein 1	CVII
DPP4	P27487	Dipeptidyl peptidase 4	CM
EFEMP1	Q12805	EGF-containing fibulin-like extracellular matrix protein 1	CM
ENG	P17813	Endoglin	CM
F11	P03951	Coagulation factor XI	CM
F7	P08709	Coagulation factor VII	CM
FABP2	P12104	Fatty acid-binding protein, intestinal	CVII

FAP*	Q12884	Prolyl endopeptidase FAP	CM
FCGR2A	P12318	Low affinity immunoglobulin gamma Fc region receptor II-a	CM
FCGR3B	O75015	Low affinity immunoglobulin gamma Fc region receptor III-B	CM
FCN2	Q15485	Ficolin-2	CM
FETUB	Q9UGM5	Fetuin-B	CM
FGF-21	Q9NSA1	Fibroblast growth factor 21	CVII
FGF-23	Q9GZV9	Fibroblast growth factor 23	CVII
FS	P19883	Follistatin	CVII
Gal-9	O00182	Galectin-9	CVII
GAS6	Q14393	Growth arrest-specific protein 6	CM
GDF-2	Q9UK05	Growth/differentiation factor 2	CVII
GH	P01241	Growth hormone	CVII
GIF	P27352	Gastric intrinsic factor	CVII
GLO1	Q04760	Lactoyl glutathione lyase	CVII
GNLY	P22749	Granulysin	CM
GP1BA	P07359	Platelet glycoprotein Ib alpha chain	CM
GT	P51161	Gastrotropin	CVII
HAOX1	Q9UJM8	Hydroxyacid oxidase 1	CVII
HB-EGF	Q99075	Proheparin-binding EGF-like growth factor	CVII
HO-1	P09601	Heme oxygenase 1	CVII
hOSCAR	Q8IYS5	Osteoclast-associated immunoglobulin-like receptor	CVII
HSP 27	P04792	Heat shock 27 kDa protein	CVII
ICAM1	P05362	Intercellular adhesion molecule 1	CM
ICAM3	P32942	Intercellular adhesion molecule 3	CM
IDUA	P35475	Alpha-L-iduronidase	CVII
IGFBP3	P17936	Insulin-like growth factor-binding protein 3	CM
IGFBP6	P24592	Insulin-like growth factor-binding protein 6	CM
IGLC2	P0DOY2	Ig lambda-2 chain C regions	CM

IL-17D	Q8TAD2	Interleukin-17D	CVII
IL-18	Q14116	Interleukin-18	CVII
IL-1ra	P18510	Interleukin-1 receptor antagonist protein	CVII
IL-27	Q14213	Interleukin-27	CVII
IL-4RA	P24394	Interleukin-4 receptor subunit alpha	CVII
IL16	Q14005	Pro-interleukin-16	CVII
IL1RL2	Q9HB29	Interleukin-1 receptor-like 2	CVII
IL6	P05231	Interleukin-6	CVII
IL7R	P16871	Interleukin-7 receptor subunit alpha	CM
ITGAM*	P11215	Integrin alpha-M	CM
ITGB1BP2	Q9UKP3	Melusin	CVII
KIM1	Q96D42	Kidney Injury Molecule	CVII
KIT	P10721	Mast/stem cell growth factor receptor Kit	CM
LCN2*	P80188	Neutrophil gelatinase-associated lipocalin	CM
LEP	P41159	Leptin	CVII
LILRB1	Q8NHL6	Leukocyte immunoglobulin-like receptor subfamily B member 1	CM
LILRB2	Q8N423	Leukocyte immunoglobulin-like receptor subfamily B member 2	CM
LILRB5	O75023	Leukocyte immunoglobulin-like receptor subfamily B member 5	CM
FCG2B	P31994	Low affinity immunoglobulin gamma Fc region receptor II-b	CVII
LOX-1	P78380	Lectin-like oxidized LDL receptor 1	CVII
LPL	P06858	Lipoprotein lipase	CVII
LTBP2*	Q14767	Latent-transforming growth factor beta-binding protein 2	CM
LYVE1	Q9Y5Y7	Lymphatic vessel endothelial hyaluronic acid receptor 1	CM
MARCO	Q9UEW3	Macrophage receptor MARCO	CVII
MBL2	P11226	Mannose-binding protein C	CM
MEGF9	Q9H1U4	Multiple epidermal growth factor-like domains protein 9	CM
MERTK	Q12866	Tyrosine-protein kinase Mer	CVII
MET	P08581	Hepatocyte growth factor receptor	CM

MFAP5	Q13361	Microfibrillar-associated protein 5	CM
MMP-12	P39900	Matrix metalloproteinase-12	CVII
MMP-7	P09237	Matrix metalloproteinase-7	CVII
NCAM1	P13591	Neural cell adhesion molecule 1	CM
NEMO	Q9Y6K9	NF-kappa-B essential modulator	CVII
NID1	P14543	Nidogen-1	CM
NOTCH1	P46531	Neurogenic locus notch homolog protein 1	CM
NRP1	O14786	Neuropilin-1	CM
OSMR	Q99650	Oncostatin-M-specific receptor subunit beta	CM
PAM	P19021	Peptidyl-glycine alpha-amidating monooxygenase	CM
PAPPA	Q13219	Pappalysin-1	CVII
PAR-1	P25116	Proteinase-activated receptor 1	CVII
PARP-1	P09874	Poly [ADP-ribose] polymerase 1	CVII
PCOLCE	Q15113	Procollagen C-endopeptidase enhancer 1	CM
PD-L2	Q9BQ51	Programmed cell death 1 ligand 2	CVII
PDGFB	P01127	Platelet-derived growth factor subunit B	CVII
PGF	P49763	Placenta growth factor	CVII
PIgR	P01833	Polymeric immunoglobulin receptor	CVII
PLA2G7*	Q13093	Platelet-activating factor acetyl hydrolase	CM
PLTP*	P55058	Phospholipid transfer protein	CM
PLXNB2	O15031	Plexin-B2	CM
PRCP*	P42785	Lysosomal Pro-X carboxypeptidase	CM
PRELP	P51888	Prolargin	CVII
PROC	P04070	Vitamin K-dependent protein C	CM
BOC	Q9BWV1	Brother of CDO	CVII
PRSS2	P07478	Trypsin-2	CM
PRSS27	Q9BQR3	Serine protease 27	CVII
PRSS8	Q16651	Prostasin	CVII

PSGL-1	Q14242	P-selectin glycoprotein ligand 1	CVII
PTPS	Q13332	Receptor-type tyrosine-protein phosphatase S	CM
PTX3	P26022	Pentraxin-related protein PTX3	CVII
QPCT	Q16769	Glutaminyl-peptide cyclotransferase	CM
RAGE	Q15109	Receptor for advanced glycosylation end products	CVII
REG1A	P05451	Lithostathine-1-alpha	CM
REG3A	Q06141	Regenerating islet-derived protein 3-alpha	CM
REN	P00797	Renin	CVII
SAA4	P35542	Serum amyloid A-4 protein	CM
SCF	P21583	Stem cell factor	CVII
SELL	P14151	L-selectin	CM
SERPINA12	Q8IW75	Serpin A12	CVII
SERPINA5	P05154	Plasma serine protease inhibitor	CM
SERPINA7	P05543	Thyroxine-binding globulin	CM
SLAMF7	Q9NQ25	SLAM family member 7	CVII
SOD1	P00441	Superoxide dismutase	CM
SOD2	P04179	Superoxide dismutase [Mn], mitochondrial	CVII
SORT1	Q99523	Sortilin	CVII
SPARCL1	Q14515	SPARC-like protein 1	CM
SPON2	Q9BUD6	Spondin-2	CVII
SRC	P12931	Proto-oncogene tyrosine-protein kinase Src	CVII
ST6GAL1	P15907	Beta-galactoside alpha-2,6-sialyltransferase 1	CM
STK4	Q13043	Serine/threonine-protein kinase 4	CVII
TCN2	P20062	Transcobalamin-2	CM
TF	P13726	Tissue factor	CVII
TGFBI	Q15582	Transforming growth factor-beta-induced protein ig-h3	CM
TGFBR3	Q03167	Transforming growth factor beta receptor type 3	CM
TGM2	P21980	Protein-glutamine gamma-glutamyltransferase 2	CVII

THBS2	P35442	Thrombospondin-2	CVII
THBS4	P35443	Thrombospondin-4	CM
THPO	P40225	Thrombopoietin	CVII
TIE1	P35590	Tyrosine-protein kinase receptor Tie-1	CM
TIE2	Q02763	Angiopoietin-1 receptor	CVII
TIMD4	Q96H15	T-cell immunoglobulin and mucin domain-containing protein 4	CM
TIMP1	P01033	Metalloproteinase inhibitor 1	CM
TM	P07204	Thrombomodulin	CVII
TNC	P24821	Tenascin	CM
TNFRSF10A	O00220	Tumour necrosis factor receptor superfamily member 10A	CVII
TNFRSF11A	Q9Y6Q6	Tumour necrosis factor receptor superfamily member 11A	CVII
TNFRSF13B	O14836	Tumour necrosis factor receptor superfamily member 13B	CVII
TNXB	P22105	Tenascin-X	CM
TRAIL-R2	O14763	TNF-related apoptosis-inducing ligand receptor 2	CVII
UMOD	P07911	Uromodulin	CM
VASN	Q6EMK4	Vasorin	CM
VCAM1	P19320	Vascular cell adhesion protein 1	CM
VEGFD	O43915	Vascular endothelial growth factor D	CVII
VSIG2	Q96IQ7	V-set and immunoglobulin domain-containing protein 2	CVII
XCL1	P47992	Lymphotactin	CVII

*Proteins not analysed because they did not reach level of assay detection.