Chronic kidney disease of undetermined cause (CKDu): an exploratory study analysing associations between an emerging epidemic and environmental risk factors

Sophie Annalisa Hamilton

Department of Epidemiology and Biostatistics
School of Public Health
Imperial College London

Thesis submitted in fulfilment of the requirements for the degree
Doctor of Philosophy (PhD)
August 2020

Abstract

An epidemic of chronic kidney disease is occurring in rural communities in low-and middle-income countries. Characterised by chronic impairment of kidney function, this disease is not associated with known risk factors diabetes or hypertension and occurs primarily in communities characterised by a hot climate with reliance on agricultural work. Hypothesised environmental risk factors include heat stress, heavy metals and pesticide exposure, and low altitude. As it is unclear whether there is a unified aetiology, this disease has been termed chronic kidney disease of undetermined cause (CKDu).

In this thesis I aimed to investigate whether the above environmental exposures are CKDu risk factors. Firstly, I conducted a global geographical analysis of CKDu using the global burden of disease database, modelling age-standardized death rates to highlight potential high-risk regions. To analyse CKDu at the individual level I developed a methodological framework to analyse satellite imagery and measure potable water sources and blood for heavy metals. I use Bayesian kernel machine regression (BKMR) to investigate associations between kidney function and heavy metals in blood. I analysed data from Sri Lanka and India, which are CKDu endemic, but aetiologies are unknown. Using satellite-derived data, associations between kidney function temperature, altitude, and vicinity to cropland were investigated. I then analysed data in Malawi which is suspected as high-risk based on similarities of climate and agriculture typical of endemic regions. Via a nested cross-sectional study, associations between kidney function and satellite-derived data were conducted. Heavy metals concentrations in potable water and blood samples were analysed using BKMR.

Results suggest CKDu is not present in Malawi. There were no associations between temperature, altitude, heavy metals, or vicinity to cropland and reduced kidney function. BKMR showed potential interactions between reduced kidney function and low blood Cadmium concentrations. In Sri Lanka and India, elevated temperature was not associated with CKDu. However, in Sri Lanka associations were observed with agricultural work; in India vicinity to cropland was associated. My findings challenge the current hypothesis that CKDu is a heat-related disease. My environmental exposure framework can be used as a foundation for further research facilitating international comparisons of environmental risk factors for CKDu.

Table of Contents

Abstract 1
List of tables 5
List of figures 9
List of equations 15
List of abbreviations 16
List of units and elements 18
Acknowledgments 20
1. Introduction 22
1.1 Chronic Kidney disease of undetermined cause 22
1.2 Aims and objectives 25
1.3 Layout of the thesis 26
2. Background 32
2.1 Structure and function of the kidney 32
2.1.2 Chronic Kidney Disease (CKD) 34
2.1.3 Chronic kidney disease of undetermined cause (CKDu) 35
2.1.4 Challenges of characterizing and diagnosing CKDu 36
2.1.5 Prevalence of CKDu 37
2.1.6 Leading hypotheses 37
2.1.6.1 Heat stress 38
2.1.6.2 Exposure to pesticides 40
2.1.6.3 Water contamination 41
2.1.6.4 Altitude 43
2.1.7 The need for standardization – the DEGREE study 50
2.1.8 Summary 51
3. Passive detection of CKDu using the Global Burden of Disease data 52
3.1 Introduction 52
3.2 Methods: 57
3.2.1 The passive detection method – an overview 57
3.2.2 Data sources 58
3.2.3 Statistical analyses 59
3.3 Results 62
3.3.1 Global spatial patterns of SMR 62
3.3.2 High income countries 62

3.3.3 Low income countries 96
3.4 Discussion 130
3.4.1 Rate ratios 136
4. Geographical analyses of environmental risk factors for CKDu across Sri Lanka and India 139
4.1 Introduction 139
4.2 Methods 141
4.2.1 Study settings, participants, study designs and health data 141
4.3 Results 160
4.3.1 Sri Lanka 160
4.3.2 India 187
4.3.2.2 Environmental characteristics 188
4.3.2.3 Mean eGFR and prevalence of low eGFR 196
4.3.2.4 Risk factors for reduced eGFR and eGFR<60 using heat index derived from empirical Bayesian kriging 196
4.3.2.5 Risk factors for reduced eGFR and eGFR<60 using heat index derived from bilinear interpolation 202
4.3.2.6 Multiple regression models 207
4.3.2.7 Linear mixed models 208
4.4 Discussion 211
4.4.1 Conclusions 215
5. A cross-sectional DEGREE study in urban and rural Malawi 217
5.1 Introduction 217
5.2 Methods 218
5.2.1 Study design and objectives 218
5.3 Results 230
5.3.1 Study population characteristics 230
5.3.2 Mean eGFR and prevalence of reduced eGFR 234
5.3.3 Risk factors for reduced eGFR and eGFR <90 235
5.3.4 Urban/rural comparison of eGFR 237
5.4 Discussion 239
5.4.1 Conclusions 242
6. Nested environmental study, Malawi 243
6.1 Introduction 243
6.2 Methods 244
6.2.1 Environmental study part 1: questionnaire and bio sample collection 244
6.2.2 Environmental study part 2: Potable water collection 251

6.2.3 Environmental study part 3: Satellite imagery assignment 259
6.3 Results 263
6.3.1 Study population characteristics 263
6.3.2 Socioeconomic and environmental characteristics 264
6.3.3 Environmental variables 264
6.3.4 Mean eGFR and prevalence of reduced eGFR 272
6.3.5 BKMR analyses: blood 276
6.3.6 Water source results 288
6.3.7 Distribution of heavy metals and trace metals 289
6.3.8 Statistical significance tests for different sampling groups 290
6.3.9 Multidimensional measures of heavy metals 291
6.3.10 Blood and water metals combined 293
6.3.11 Modelling the harmonized dataset 297
6.3.1.2 Modelling data using bilinear interpolation heat index data 303
6.4 Discussion 308
7 General discussion 313
7.3 Summary of findings and their public health implications 313
7.4 An overview of the findings in this thesis 315
7.5 Public health implications of CKDu 316
7.6 Passive detection 318
7.7 Heat index and altitude 319
7.8 Vicinity to crop land, occupation, and self-reported symptoms 321
7.9 Water Sources 323
7.10 Heavy metals in potable water 324
7.11 Heavy metals biomarkers 325
7.12 The issue of causality in CKDu research 326
7.13 Implications of CKDu at the individual level 328
7.14 Academic research implications 328
7.15 Key further research questions 329
7.16 Conclusions 331
Bibliography 332

List of tables

Table 1. Summary of data sources and methods used for each chapter in this thesis 28

Table 2. Summary of key papers used in the literature review 46

Table 3. Table showing the 10th revision of the standardized international classification of diseases (ICD-10) codes, and respective Global Burden of Disease re-coding categories used to estimate the SMR values for: "all cause CKD", "CKD attributed to diabetes" and "non-diabetes related CKD" 59

Table 4. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: "all cause CKD", "CKD attributed to DM" and "non-DM related CKD" for males 15-49 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 63

Table 5. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: "all cause CKD", "CKD attributed to DM" and "non DM related CKD" for males 50-69 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 68

Table 6. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death :"all cause CKD", "CKD attributed to DM" and "non DM related CKD" for males 70+ years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 74

Table 7. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death for "all cause CKD", "CKD attributed to DM" and "non DM related CKD" for females 15-49 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 79

Table 8. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death :"all cause CKD", "CKD attributed to DM" and "non DM related CKD" for females 50-59 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 85

Table 9. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death:"all cause CKD", "CKD attributed to DM" and "non DM related CKD" for females 70+ years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 91

Table 10. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death :"all cause CKD", "CKD attributed to DM" and "non DM related CKD" for males 15-49 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 97

Table 11. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: "all cause CKD", "CKD attributed to DM" and "non-DM related CKD" for males 50-69 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 102

Table 12. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death "all cause CKD", "CKD attributed to diabetes" and "non diabetes related CKD" for males 70+ years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 108

Table 13. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: “all cause CKD”, “CKD attributed to diabetes” and “non diabetes related CKD” for females 15-49 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 113

Table 14. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: “all cause CKD”, “CKD attributed to diabetes” and “non diabetes related CKD” for females 50-69 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 119

Table 15. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: “all cause CKD”, “CKD attributed to diabetes” and “non diabetes related CKD” for females 70+ years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark..... 125

Table 16. Design and recruitment methods for the studies used in the India CKDu analysis..... 147

Table 17. Summary of satellite-derived environmental data sources used in this analysis..... 148

Table 18. Empirical Bayesian Kriging cross-validation summary for heat index across Sri Lanka 161

Table 19. K-fold cross validation Summary of the performance of interpolation methods to predict heat index across Sri Lanka..... 166

Table 20. Correlation analysis of interpolation methods across Sri Lanka 167

Table 21. Associations of sociodemographic and anthropometric characteristics with low eGFR in participants without diabetes, hypertension and heavy proteinuria using Empirical Bayesian kriging derived heat index data in Sri Lanka, n=3297..... 170

Table 22. Associations of sociodemographic and anthropometric characteristics with eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in Sri Lanka modelled using Empirical Bayesian kriging derived heat index data, n=3297 172

Table 23. Associations of sociodemographic and anthropometric characteristics with low eGFR in participants without diabetes, hypertension and heavy proteinuria using bilinear interpolation derived heat index data in Sri Lanka, n=3297..... 175

Table 24. Associations of sociodemographic and anthropometric characteristics with eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in Sri Lanka modelled using bilinear interpolation derived heat index data, n=3297..... 178

Table 25. Fully adjusted logistic regression by study area in Sri Lanka (n=3297) 181

Table 26. Multiple regression Sri Lanka heat index analyses comparing EBK method and bilinear interpolation data with altitude coefficients appended (n=3297) 184

Table 27. Fixed effects table of environmental factor effect (heat index, altitude, land cover, pesticide use, chena cultivation, rice and vegetable farming and water source) on eGFR across zones 1-3. 187

Table 28. Empirical Bayesian Kriging cross validation summary for UDAY and CARRS time periods... 189

Table 29. K fold cross validation Summary of the performance of interpolation methods to predict heat index across India..... 190

Table 30. Correlation analysis of interpolation methods across India..... 190

Table 31. Associations of sociodemographic, anthropometric, and environmental characteristics with eGFR and eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in India including heat index via Empirical Bayesian kriging, n=11,119..... 198

Table 32. Associations of sociodemographic and anthropometric characteristics with eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in Sri Lanka modelled with Empirical Bayesian Kriging heat data (n=11,119) 200

Table 33. Associations of sociodemographic, anthropometric, and environmental characteristics with eGFR in participants without diabetes, hypertension and heavy proteinuria in India including heat index via bilinear interpolation, n=11,119 203

Table 34 Associations of sociodemographic and anthropometric characteristics with eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in Sri Lanka modelled with bilinear interpolation heat data (n=11,119) 205

Table 35. Multiple regression models of sociodemographic and environmental characteristics associated with eGFR and eGFR<60 India 207

Table 36. Linear mixed model for altitude, heat index and land cover in India 208

Table 37. Fixed effects table of environmental factor (heat index, Land cover and altitude) effect on eGFR across Indian and Sri Lankan study zones 211

Table 38. Sociodemographic and anthropometric characteristics of study participants without diabetes, hypertension and heavy proteinuria) n=821 232

Table 39. Associations between sociodemographic and anthropometric characteristics and eGFR and eGFR<90 in participants without diabetes, hypertension and heavy proteinuria, n=821 236

Table 40. Associations between sociodemographic and anthropometric characteristics and estimated glomerular filtration rate (eGFR) (fully adjusted) after removal of those with hypertension, diabetes and proteinuria, Area 25 (n=243) and Bonje (n=578) 238

Table 41. Summary of variables collected in the environmental module questionnaire as art of data collection for the nested study 246

Table 42. Water source parameters measured in the field and apparatus used 256

Table 43. Summary of satellite-derived environmental data sources used in environmental analyses for Bonje and Area 25, Malawi 259

Table 44. Empirical Bayesian Kriging hold out validation summary for heat index across Malawi 265

Table 45. K-fold cross validation summary of the performance of interpolation methods to predict heat index across Malawi 271

Table 46. Correlation analysis of interpolation methods across Malawi 272

Table 47. Socio-economic and demographic characteristics of study participants in the Malawian sub-sample (n=367) 273

Table 48. Descriptive blood metals results in Bonje and Area 25 276

Table 49. Multivariable liner regression model of eGFR and heavy metals blood concentrations stratified by study zone in Malawi 277

Table 50. Descriptive statistics of physicochemical water parameters by water source in Bonje and Area 25 288

Table 51. Summary of heavy metals concentrations across all water sources in Bonje 289

Table 52. Summary of heavy metals concentrations across all water sources in Area 25 290

Table 53. Kruskal Wallis test for heavy metals and pairwise associations of heavy metals concentrations between different water sources (µg/L) within and across Bonje and Area 25..... 291

Table 54. Eigenvalues and associated indicators for PCA analysis on heavy metals..... 293

Table 55. T test of varimax-rotated principal component (RC) scores to assess the mean difference in heavy metals concentrations between Bonje and Area 25 across all water sources 293

Table 56. Linear regression model of metals in blood and water and water source in association with eGFR 297

Table 57. Associations between sociodemographic, anthropometric, and environmental characteristics and eGFR in Malawian participants using heat index data from empirical Bayesian kriging, n=367..... 299

Table 58. Associations between sociodemographic, anthropometric, and environmental characteristics and eGFR <90 in Malawian participants using heat index data from empirical Bayesian kriging, n=367..... 301

Table 59. Associations between sociodemographic, anthropometric, and environmental characteristics and eGFR in Malawian participants using heat index data from bilinear interpolation methodology, n=367 304

Table 60. Associations between sociodemographic, anthropometric, and environmental characteristics and eGFR <90 in Malawian participants using heat index data from bilinear interpolation methodology, n=367 306

Table 61. Time and financial cost for setting up and rolling out the environmental module framework in other DEGREE sites..... 329

List of figures

Figure 1. Diagram of the thesis structure showing the "funnel" approach of investigating CKDu from the global level using the Global Burden of Disease (GBD) data, followed by investigation of CKDu in countries which have reported the disease but lack environmental analyses, followed by a cross sectional and nested study of CKDu in Malawi which is suspected to be at risk of the disease 27

Figure 2. Simplified structure of the kidney (image reproduced from Encyclopaedia Britannica) 32

Figure 3. Structure of the glomerulus and Bowman's Capsule 33

Figure 4. Classification of CKD prognosis by eGFR categories and albuminuria levels as classified by the Kidney Disease Improving Global Outcomes (KDIGO). Risk colour spectrum shows low risk (green); moderately increased risk (yellow); high risk (orange) and very high risk (red) categories 35

Figure 5. Map showing data quality of cause of death according to the World Health Organization Global Health Estimates Project 54

Figure 6. Global prevalence of treated end stage renal failure (ESRD) (per 1^6 population) based on individual country data. Source: Bello et.al, 2017 57

Figure 7. Global distribution of World Bank income groups by country 61

Figure 8a. Spatial distribution of SMR values for all-cause CKD in 15-49-year-old males in higher income countries. Inserts highlight smaller countries 64

Figure 8b. Spatial distribution of SMR values for CKD attributed to diabetes for 15-49-year-old males in higher income countries. Insert highlights smaller countries 65

Figure 8c. Spatial distribution of SMR values for non-diabetes related CKD for 15-49-year-old males in higher income countries. Inserts highlight smaller countries 66

Figure 8d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 15-49-years-old in higher income countries 67

Figure 9a. Spatial distribution of SMR values for all-cause CKD for 50-69-year-old males in higher income countries. Inserts highlight smaller countries 70

Figure 9b. Spatial distribution of SMR values for CKD attributed to diabetes CKD for 50-69-year-old males in higher income countries. Inserts highlight smaller countries 71

Figure 9c. Spatial distribution of SMR values for Non-diabetes related CKD for 50-69-year-old males in higher income countries. Inserts highlight smaller countries 72

Figure 9d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 50-69-year-old in higher income countries 73

Figure 10a. Spatial distribution of SMR values for all-cause CKD for 70+ year males in higher income countries. Inserts highlight smaller countries 75

Figure 10b. Spatial distribution of SMR values for CKD attributed to diabetes CKD for 70+ year males in higher income countries. Inserts highlight smaller countries 76

Figure 10c. Spatial distribution of SMR values for non-diabetes related CKD for 70+ year males in higher income countries. Inserts highlight smaller countries 77

Figure 10d. Spatial distribution of the Mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 70+ years in higher income countries 78

Figure 11a. Spatial distribution of SMR values for all-cause CKD for 15-49-year females in higher income countries. Inserts highlight smaller countries 81

Figure 11b. Spatial distribution of SMR values for CKD attributed to diabetes for 15-49-year females in higher income countries. Inserts highlight smaller countries 82

Figure 11c. Spatial distribution of SMR values for non-diabetes related CKD for 15-49-year females in higher income countries. Inserts highlight smaller countries 83

Figure 11d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for females 15-49 years in high income countries 84

Figure 12a. Spatial distribution of SMR values for all-cause CKD for 50-69-year females in higher income countries. Inserts highlight smaller countries 87

Figure 12b. Spatial distribution of SMR values for CKD attributed to diabetes for 50-69-year females in higher income countries. Inserts highlight smaller countries 88

Figure 12c. Spatial distribution of SMR values for non-diabetes related CKD for 50-69-year females in higher income countries. Inserts highlight smaller countries 89

Figure 12d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for females 50-69 years in higher income countries. Inserts highlight smaller countries . 90

Figure 13a. Spatial distribution of SMR values for all-cause CKD for 70+ year females in higher income countries. Inserts highlight smaller countries 92

Figure 13b. Spatial distribution of SMR values for CKD attributed to diabetes for 70+ year females in higher income countries. Inserts highlight smaller countries 93

Figure 13c. Spatial distribution of SMR values for non-diabetes related CKD for 70+ year females in higher income countries. Inserts highlight smaller countries 94

Figure 13d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for females 70+ years in higher income countries. Inserts highlight smaller countries 95

Figure 14a. Spatial distribution of SMR values for all-cause CKD for 15-49-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 98

Figure 14b. Spatial distribution of SMR values for CKD attributed to diabetes for 15-59-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 99

Figure 14c. Spatial distribution of SMR values for non-diabetes related CKD for 15-49-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 100

Figure 14d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 15-49-year-old in lower income countries 101

Figure 15a. Spatial distribution of SMR values for all-cause CKD for 50-69-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 104

Figure 15b. Spatial distribution of SMR values for CKD attributed to diabetes CKD for 50-69-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 105

Figure 15c. Spatial distribution of SMR values for non-diabetes related CKD for 50-69-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 106

Figure 15d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 50-69-year-old in lower income countries 107

Figure 16a. Spatial distribution of SMR values for all-cause CKD for 70+ year males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 109

Figure 16b. Spatial distribution of SMR values for CKD attributed to diabetes for 70+ year males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 110

Figure 16c. Spatial distribution of SMR values for non-diabetes related CKD for 70+ year males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 111

Figure 16d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 70+ years in lower income countries..... 112

Figure 17a. Spatial distribution of SMR values for all-cause CKD for 15-49-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 115

Figure 17b. Spatial distribution of SMR values for CKD attributed to diabetes for 15-49-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia..... 116

Figure 17c. Spatial distribution of SMR values for non-diabetes related CKD for 15-49-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 117

Figure 17d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for 15-49-year females in lower income countries 118

Figure 18a. Spatial distribution of SMR values for all-cause CKD for 50-69-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 121

Figure 18b. Spatial distribution of SMR values for CKD attributed to diabetes for 50-69-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia..... 122

Figure 18c. Spatial distribution of SMR values for non-diabetes-related CKD for 50-69-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 123

Figure 18d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related CKD for females 50-69 years in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 124

Figure 19a. Spatial distribution of SMR values for all-cause CKD for 70+ year females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 126

Figure 19b. Spatial distribution of SMR values for CKD attributed to diabetes for 70+ year females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 127

Figure 19c. Spatial distribution of SMR values for non-diabetes-related CKD for 70+ year females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 128

Figure 19d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for females 70+ years in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia 129

Figure 20. Map of Sri Lanka showing the five communities selected for the DEGREE survey in the North Central Province, Sri Lanka. 142

Figure 21. Map of Sri Lanka DEGREE study zones grouped according to governmental classification of CKD/u endemicity. 143

Figure 22. Map of India showing the "*Implementing a Comprehensive Diabetes Prevention Management Program"* (UDAY) study sites located in rural and urban Sonipat (Haryana) and Vishakhapatnam (Andhra Pradesh). 145

Figure 23. Map of India showing the "*Centre for Cardiometabolic Risk Reduction in South Asia"* (CARRS) study sites located in urban Delhi (Delhi) and rural Chennai (Tamil Nadu). 146

Figure 24: Diagram of a Semivariogram showing the main components (i) nugget (ii) range and (iii) sill. X axis shows the distance (lag) from the measuring point, Y axis shows average variogram value Source: adapted from ESRI [(242)] 151

Figure 25: Spectrum of simulated semivariogram models produced using the Empirical Bayesian kriging method. X axis shows the distance (lag) from the measuring point, Y axis shows average variogram value 152

Figure 26. Infographic describing the resampling method for bilinear interpolation. (Source ESRI[(248)]). 155

In order to understand whether the choice of model affects the covariates in the regression models, I conducted modelling with the EBK heat index data and the bilinear interpolation heat index data with the aim of highlighting the importance of selecting the most appropriate model for predictive surfaces in environmental exposure analyses. 155

Figure 27. Study flowchart with exclusion criteria for the Sri Lanka DEGREE population sample 160

Figure 28. Environmental variable surfaces across Sri Lanka DEGREE study zones showing (a) Heat index, (b) Land cover (see appendix 4 for entire legend), (c) Altitude 163

Figure 29. Environmental variable surfaces across Sri Lanka showing (a) Heat index, (b) Land cover (see appendix 4 for entire legend), (c) Altitude 164

Figure 30. Prediction standard error maps for Empirical Bayesian Kriging (EBK) interpolation of heat index data across (a) India for CARRS study period (b) India for UDAY study period (c) Sri Lanka 165

Figure 31. Distribution of estimated glomerular filtration rate (eGFR) levels by sex and study zone in Sri Lanka 168

Figure 32. Linear mixed model caterpillar plot of eGFR in study zones 1-3 in Sri Lanka. Intercept denotes overall mean eGFR across the study zones; blue circles represent the average eGFR for each zone; black bars represent 95% confidence intervals. 186

Figure 33. Study flowchart with exclusion criteria for the India population sample 188

Figure 34. Environmental variable surfaces across Andhra Pradesh study zones showing (a) Heat index, (b) land cover, (c) altitude 191

Figure 35. Environmental variable surfaces across Delhi study zone showing (a) Heat index, (b) land cover, (c) altitude 192

Figure 36. Environmental variable surfaces across Sonipat study zones showing (a) Heat index, (b) land cover, (c) altitude 193

Figure 37. Environmental variable surfaces across Andhra Pradesh study zones showing (a) Heat index, (b) land cover, (c) altitude 194

Figure 38. Environmental variable surfaces across India study zones showing (a) Heat index, (b) Altitude, (c) Land cover 195

Figure 39. eGFR by sex and state in India 196

Figure 40. Linear mixed model caterpillar plot of eGFR in study zones in India accounting for land cover, heat index and altitude. Intercept denotes the overall mean eGFR across the study zones; blue circles represent the average eGFR for each zone; black bars represent 95% confidence intervals... 208

Figure 42. Map showing the urban and rural DEGREE study zones in Malawi 220

Figure 43. Map showing the Health and Demographic Surveillance Site (HDSS) study zones numbered 1-21, and colour coded to denote zone boundary 221

Figure 44. Map showing the Area 25 study zones 1-35, which are colour coded to denote zone boundary 222

Figure 45. Flowchart showing the nested sample frames from which the DEGREE participants in Karonga (Bonje) and Lilongwe (Area 25) were sampled 224

Figure 46. DEGREE exclusion criteria used for the Malawian sample frame, and final sample size used for analysis. 231

Figure 47. Histogram of eGFR distribution in the sample population across urban and rural Malawi 235

Figure 48. Photos of frequently used water sources in Bonje showing (a) a well; and (b) a borehole with an Afridev pump system 251

Figure 49. Photos of frequently used water sources in Area 25 showing (a) a water kiosk; and (b) a well 252

Figure 50. Map showing the location of Bonje village within the HDSS, and the location of Hara Rice Scheme 253

Figure 51. Maps showing the sample frame buffers applied to each study zone for water collection in Bonje and Area 25 255

Figure 52. Water sources sampled in (a) Bonje and (b) Area 25. Sources are numbered according to the order in which they were sampled 258

Figure 53. Study flowchart with exclusion criteria for the Malawi nested study population sample (n=367) 264

Figure 54. Land cover (a), heat index (b) and altitude (c) surfaces of Malawi modelled using Empirical Bayesian Kriging; refer to figure 40 for study zone locations 267

Figure 55. Land cover (a), heat index (b) and altitude (c) surfaces of Bonje modelled using Empirical Bayesian Kriging 268

Figure 56. Land cover (a), heat index (b) and altitude (c) surfaces of Area 25 modelled using Empirical Bayesian Kriging 269

Figure 57. Prediction standard error maps for Empirical Bayesian Kriging interpolation of heat index data across Malawi 270

Figure 58. Histogram of eGFR distribution in the nested sample population in Malawi (n=367) 273

Figure 59. Overall effect of the blood heavy metals mixture on eGFR (y axis) in Bonje when metals are at a specific percentile (shown on the x axis) in comparison to then exposures are all held at the 50th percentile 279

Figure 60. Single metal association on eGFR (x axis) for each metal at its 75th versus 25th percentile, when all other metals are held at either the 25th (red), 50th (green) or 75th (blue) percentile (y axis) in Bonje. Plot shows estimates (dots) and 95% credible intervals (bars), the grey dotted line indicates the null 280

Figure 61. Univariate dose-response functions of heavy metals (x axis) eGFR (y axis) in Bonje. Plot shows 95% confidence intervals for each metal, with all other metals fixed at the 50th percentile 281

Figure 62. Interactions between metals on eGFR in Bonje. Bivariate exposure-response functions for the metal on the x axis when the metal on the y axis is either at the 25th (red), 50th (green), or 75th (blue) percentile 282

Figure 63. Overall effect of the blood heavy metals mixture on eGFR (y axis) in Area 25 when exposures are at a specific percentile (shown on the x axis) in comparison to then exposures are all held at the 50th percentile 284

Figure 64. Single metal association on eGFR (x axis) for each metal at its 75th versus 25th percentile, when all other metals are held at either the 25th (red), 50th (green) or 75th (blue) (y axis) in Area 25. Plot shows estimates (dots) and 95% credible intervals (bars), the grey dotted line indicates the null 285

Figure 65. Univariate exposure-response functions of heavy metals (x axis) on eGFR (y axis) in Area 25. Plot shows 95% confidence intervals for each metal, with all other metals fixed at the 50th percentile 286

Figure 66. Interactions between metals on eGFR in Area 25. Bivariate exposure-response functions for the metal on the x axis when the metal on the y axis is either at the 25th (red), 50th (green), or 75th (blue) percentile 287

Figure 67. Scree plot showing the cumulative variance explained (y axis) by each principal component (x axis) 292

Figure 68. Heat map showing the correlation of heavy metals in water sources across Bonje and Area 25 292

Figure 69. Water sources assigned to participants using (a) 500m buffers across Bonje and (b) 200m buffers across Area 25 294

Figure 70. Correlation matrix of blood and water heavy metals in Bonje. Metals labelled "_h2o" denote water metals and those without denote blood metals. Narrower ellipses indicate stronger correlations, and the ellipses' orientation indicates the correlation direction. Cells display significant correlations ($p<0.05$) 295

Figure 71. Correlation matrix of blood and water heavy metals in Area 25. Metals labelled "_h2o" denote water metals and those without denote blood metals. Narrower ellipses indicate stronger correlations, and the ellipses' orientation indicates the correlation direction. Cells display significant correlations ($p<0.05$) 296

List of equations

Equation 1. SMR formula 60
Equation 2. Expected deaths formula 60
Equation 3. Heat index formula 149
Equation 4. Variogram model equation 150
Equation 5. Log transformation adding a constant to negative heat index values 153
Equation 6. Inverse distance weighting equation 154
Equation 7. CKD-EPI equation to calculate eGFR 228
Equation 8. BKMR formula 262

List of abbreviations

2,4-D	2,4-Dichlorophenoxyacetic acid
4V-MDRD	Four-variable Modification of Diet in Renal Disease
ACR	Albumin: Creatinine ratio
AKI	Acute kidney injury
ART	Antiretroviral Therapy
BDL	Below detectable limits
BIA	Bioimpedance analysis
BKMR	Bayesian kernel machine regression
BMI	Body Mass Index
CARRS	Centre for cardiometabolic Risk Reduction in South-Asia
CI	Confidence interval
CKD	Chronic Kidney Disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CKDu	Chronic Kidney Disease of undetermined cause
CRR	Caribbean Renal Registry
DEFRA	Department for Environment, Food and Rural Affairs
DEGREE	Disadvantaged Populations eGFR Epidemiology
DEM	Digital Elevation Model
DM	Diabetes mellitus
DO	Dissolved Oxygen
EA	Enumeration area
EBK	Empirical Bayesian Kriging
EC	Electrical conductivity
ECMWF	European Centre for Medium-Range Weather Forecasts
EDTA	Ethylenediamine tetraacetic acid
eGFR	Estimated glomerular filtration rate
EMQ	Environmental module questionnaire
EPO	Erythropoietin
ESA-CCI	The European Space Agency - Climate Change Initiative
ESA-S2-LC20	European Space Agency- Climate Change Initiative Land Cover Sentinel-2 Prototype-Land cover at 20 metres
ESRD	End stage renal disease
EWAS	Environment-wide association study
FFM	Fat-free mass
GBD	Global burden of disease
GIS	Geographical Information System
GPS	Global Positioning System
HARS	high altitude renal syndrome
HDSS	Health and Demographic Surveillance Site
HI	Heat index

HINTS	Health Information National Trends Study
HIV	Human immunodeficiency virus
i3C	The International Society of Nephrology's International Consortium of Collaborators on Chronic Kidney Disease of Unknown Etiology
IC	Ion chromatography
ICD-10	International Classification of Diseases - 10th edition
ICP-MS	Inductively coupled plasma mass spectrometry
IDW	Inverse distance weighting
IQR	Inter quartile range
ISN	International Society of Nephrology
KDIGO	Kidney Disease Improving Global Outcomes
KPS	Karonga Prevention Study
LMIC	Low and middle-income countries
LMM	Linear mixed model
MBS	Malawi Bureau of Standards
MCMC	Markov chain Monte Carlo algorithm
MDRF	Madras Diabetes Research Foundation
MEIRU	Malawi Epidemiology and Intervention Research Unit
MeN	Mesoamerican Nephropathy
MK	Malawian Kwatcha
MTA	Material transfer agreement
NCD	Noncommunicable disease
NCP	North Central Province
NKF	National Kidney Foundation
ODK	Open Data Kit
OR	Odds ratio
OSHA	Occupational Safety & Health Administration
PCA	Principal component analysis
PHFI	Public Health Foundation of India
PI	Principal investigator
Q-Q plot	Quantile-quantile plot
RC	Rotated component
RIQAS	Randox International Quality Assurance Scheme
RMSE	Root mean squared error
RO	Reverse osmosis
RRT	renal replacement therapy
SCr	Serum Creatinine
SD	Standard deviation
SDI	Socio-demographic Index
SMR	Standardized mortality ratio
SPECNO	Specimen number

SRTM	The Shuttle Radar Topography Mission
SSA	Sub-Saharan Africa
STID	Study identification number
TDS	Total dissolved solids
UDAY study	Implementing a comprehensive diabetes prevention and management program study
VIF	Variance inflation factor
WHO	World Health Organization

List of units and elements

^{o}C	degrees Celsius
kg/m^2	kilogram per square metre
km	kilometre
m	metre
µg/L	microgram per litre
µS/cm	micro siemens per centimetre
mg/mmol	milligram per millimole
ml	millilitre
mm Hg	millimetre of mercury
NTU	Nephelometric Turbidity Units

As	Arsenic
Al	Aluminium
Ba	Barium
Ca	Calcium
$CaCO_3$	Calcium carbonate
Cd	Cadmium
Cr	Chromium
F^-	Fluoride
HCL	Hydrochloric acid
Hg	Mercury
Pb	Lead
U	Uranium
V	Vanadium

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Declaration of originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge, it contains no material previously written by another person, except where due reference has been made in the text.

Acknowledgments

I would like to thank my supervisors Paolo Vineis, Neil Pearce and Daniela Fecht for their guidance and support in all the phases of my research. Paolo thank you for your enthusiasm and insightful discussions on my research throughout my PhD. Daniela thank you for always offering guidance on all things GIS (and otherwise) related and thank you for answering my questions – no matter how small! I would like to thank Neil for welcoming me so warmly into the DEGREE and NCD teams at LSHTM and for all your guidance throughout this process. I am truly grateful for everything you have taught me.

This thesis could not have been completed without the help and collaboration of the Malawian research team at MEIRU: Dr Crampin, Dr Prynn and Dr Nakanga. Thank you for taking me on board, teaching me about the clinical aspects of the study and letting me loose in the field! A very special thanks goes out to the fieldwork and laboratory teams in Malawi: Veronica, Cynthia, Jimmy, Fred, Odala, Oddie, Lawrence, Louis, and Marriott. You welcomed me to Malawi and educated me far beyond any books ever could. I would also like to thank Dr Scott Young and his team at the University of Nottingham who analysed my samples and gave great insight into the science of water and blood analyses.

From a personal perspective absolutely none of this could have been done without my family and friends. Dad, you are my compass. Thank you for always supporting my decisions and having endless belief in me. Thank you for the guidance, inspiration and calm you have provided throughout my life. Mum, my powerhouse and uplifter. You keep me going when I feel all else is lost. You have taught me compassion, and to always stop and appreciate the simple pleasures of life, no matter how difficult things get. Patrick thanks for all the emergency coding help when I was in the depths of the Malawian field, midway through a monsoon. Thanks for also being generally solid, and for the constant supply of artichokes! Gina thank you for being my ray of sunshine, for being my adventure buddy, and for that eagle eye! To the mighty atom Coco – thank you for making life shine again.

And last but by no means least, an all-encompassing thank you to Ben. You have been my best friend for 10 years, but over the last 4 years your support and love has been immeasurable. Hong Kong was the best decision I ever made.

I love you all beyond words.

For George and Maggie

Monty…

1. Introduction

1.1 Chronic Kidney disease of undetermined cause

An epidemic of chronic kidney disease is occurring in rural communities in an increasing number of low-and middle-income countries (LMICs)[1, 2]. Characterised by chronic impairment of kidney function, this disease is not associated with known risk factors such as diabetes, hypertension, or proteinuria (proteins in the urine), and occurs primarily in communities characterised by a hot climate with reliance on heavy agricultural work [2]. As it is currently unclear whether there is a unified underlying aetiology, this disease has been termed chronic kidney disease of undetermined cause (CKDu), and is estimated to have led to the premature deaths of tens to hundreds of thousands of young men and women over the past 2 decades[1]. Currently, the highest prevalence of CKDu has been reported in Nicaragua, Central America where 10-20% of the adult population in rural areas are affected[3, 4]. The disease has also been reported in other rural communities across Southern India and Sri Lanka where the prevalence is estimated to be 1.6% and 6% respectively [5, 6]. To date, prevalence estimates in studies conducted in different regions have been highly variable due to heterogenous study designs, sampling approaches and case definitions, which makes international comparisons a significant challenge[2].

CKDu is notoriously difficult to diagnose as it remains clinically silent for much of its course, making detection in seemingly asymptomatic individuals a challenge. Kidney function is measured using the estimated glomerular filtration rate (eGFR) which is an estimation of the volume of blood filtered through the kidney per minute (as described in chapter 2), and normal kidney function is defined by an eGFR >90 ml/min/1.73m^2 body mass area[7]. CKDu however is defined by having an eGFR of <60 ml/min/1.73m^2 for more than 3 months in the absence of diabetes, hypertension, old age, or proteinuria[1, 2, 8]. Measuring eGFR can be a costly process, and in many resource scarce CKDu endemic regions, research groups have instead used routinely collected data from renal registries to “passively detect” CKDu in place of high-cost studies [2, 9, 10]. This method has inherent limitations due to the variation in eGFR estimation practices and the clinical stratification used in different countries which can result in misclassification of CKD/u, and potentially reporting presence of a disease which might not actually exist in a population [10].

In order to improve our understanding of the aetiology, pathophysiology, and true prevalence of CKDu, reliable, standardised methodologies to actively measure eGFR in a population are needed which can then be used in future comparative CKDu studies across the globe. Coupled with the need for these standardized measurements, is the need to also elucidate the potential environmental aetiologies of CKDu. To date, environmental risk factors have been a key research stream in the CKDu literature, and several potential causative factors have been implicated. Key among these are:

(i) high ambient temperatures[11-15]

(ii) physically stressful working environments [16, 17]

(iii) low altitude [18, 19]

(iv) exposure to heavy metals via potable water [20-25]

(v) exposure to pesticides via agricultural farming practices [3, 22, 26, 27]

(vi) residential vicinity to cropland as a proxy for exposure to pesticides [10]

A major issue however is that environmental research on CKDu has been conducted in silos aiming to link the disease to a single, localized environmental risk factor, as opposed to taking a multi-factorial approach. For example, research in Central America has emphasised heat as a main aetiology for CKDu, whereas in Sri Lanka, exposure to heavy metals via ground water sources is a key research focus[14, 28-31]. In addition to the lack of multi-factorial investigations in the research field, environmental data collection at more granular detail across endemic regions is lacking. To date it has therefore remained unclear whether risk factors such as exposure to heavy metals, high ambient temperatures or living and working at low altitude are indeed associated with CKDu. In order to establish which environmental risk factors are associated with this disease, a standardized multifactorial environmental framework is essential in exploratory studies of multiple risk factors. This framework can serve as a foundation for future CKDu research to include exposure assessments, cohort, and intervention studies.

CKDu studies from Central America and South Asia suggest that rural sub-tropical populations are most likely at risk[26, 29]. These characteristics have largely influenced the locations of subsequent CKDu studies, and there is now uptake of research focus in other sub-tropical countries. Until recently little kidney function information has been available from African countries which have similar climatic and occupational characteristics to the Central American

and South Asian reporting regions. Recent research from Malawi in East Africa has demonstrated a significant association between impaired kidney function and agricultural labour, which is a central hypothesis in the CKDu literature [32]. As a result of these findings, the country is now being investigated for presence of CKDu.

In order to address the issue of eGFR standardization in current and future CKDu studies, a newly developed protocol has been suggested for global research groups. The Disadvantaged Populations eGFR Epidemiology Study (DEGREE) protocol[2] is an important cross-sectional study framework developed by clinicians and epidemiologists; this measures kidney function in at-risk communities using a standardized methodology to aid international comparisons of the disease. The scope of this protocol however does not cover individual-level environmental exposure analyses, or collection of bio samples for metals or pesticide testing which are public health concerns in many CKDu-affected populations. Without such data, it is not possible to establish the role of the current hypothesised risk factors in CKDu, and in turn allow for international comparisons. Therefore using a range of protocolled data collection methods such as questionnaires, measuring biomarkers, assessment of water sources, and satellite-derived data, a methodological framework can be constructed to objectively define the significance of environmental exposures at the individual-level in CKDu.

It must be noted that CKDu is not the first global epidemic of unexplained chronic kidney disease to puzzle the clinical and epidemiological research communities. Historical CKD epidemics of note include the "Itai-Itai" ("ouch-ouch") disease in Japan[33, 34] and Balkan endemic nephropathy[35, 36], both of which (much like CKDu) took decades of cohesive international research collaboration and adoption of multifactorial approaches to exploring biological and environmental risk factors, to discover their aetiologies.

Six decades after it was reported in 1912, it was discovered that the "*Itai-Itai"* epidemic was a Cadmium-induced nephropathy from exposure to wastewater from a mine near the Jinzu river basin in Toyama, Japan[34]. This Cadmium-induced nephropathy was characterised by interstitial fibrosis with tubular atrophy and still exists today. Cadmium remains a key risk factor for kidney damage, and a major focus of research across the globe. Balkan endemic nephropathy was first reported in rural villages along the Danube River in the 1960s and was characterized by urothelial tumours[37]. Affected villages were situated next to unaffected

ones, and familial clustering suggested possible genetic susceptibility. In the mid-1990s it was discovered that ingestion of herbal slimming remedies and flour contaminated with seeds from *Aristolochia clematitis* was the cause of this epidemic[35]. More recently, this was confirmed when *aristolactam* (AL)−DNA adducts were observed in the renal cortex of patients with urothelial tumours. From ongoing global surveillance of this epidemic, Taiwan is thought to have the world's highest prevalence of urothelial malignancy, and cases have also been reported in Australia, North America, and Europe. There is no specific treatment, so surveillance is still active, and treatment therapies are supportive, aiming to delay disease progression[37].

Much can be gleaned from the cohesive and multifactorial approaches which were taken to elucidate the aetiologies of these epidemics and to establish effective surveillance networks. In order to disentangle the mystery of CKDu and its environmental risk factors, similar transparent, and standardized approaches must be taken.

1.2 Aims and objectives

The aim of this thesis is to investigate whether currently hypothesised environmental risk factors are associated with CKDu, and whether any associations found are common across all suspected high-risk countries.

In order to investigate this key research question, I will conduct analyses in (i) Sri Lanka and India which have existing data which report the disease, but lack granular environmental data, (ii) In Malawi, which is suspected to be at risk of CKDu based on having similar environmental and socioeconomic characteristics as CKDu endemic countries. Furthermore, the issue of "passive detection" of CKDu will be explored as this could be a valuable method in developing nations where CKDu is suspected but resource limitations prohibit cross-sectional studies.

These aims will be met by:

(i) Conducting a global detection of CKDu using the Global Burden of Disease dataset by applying the "passive detection" method to assess whether use of routinely collected data sources are appropriate for CKDu detection

(ii) Undertaking exploratory environmental analyses using satellite derived imagery and climate reanalysis data in Sri Lanka and India where CKDu has been reported, to investigate whether temperature, altitude, vicinity to agricultural land and farming practices are associated with a low eGFR

(iii) Analysing the DEGREE cross-sectional study in Malawi, which is considered at risk of CKDu, to determine the distribution and prevalence of low eGFR (<60mL/min per 1.72m^2) in an urban and rural population

(iv) Using the DEGREE cross-sectional study as a sample frame for a nested environmental study in Malawi investigating individual level environmental exposures:

a. Heat
b. Altitude
c. Heavy metals in potable water and blood samples
d. Physical symptoms of heat stress resulting from occupation
e. Residential vicinity to cropland as a proxy for pesticide exposure

1.3 Layout of the thesis

This thesis follows a funnel-like structure (see Figure 1), whereby I investigate the potential risk factors for CKDu in different regions of the world.

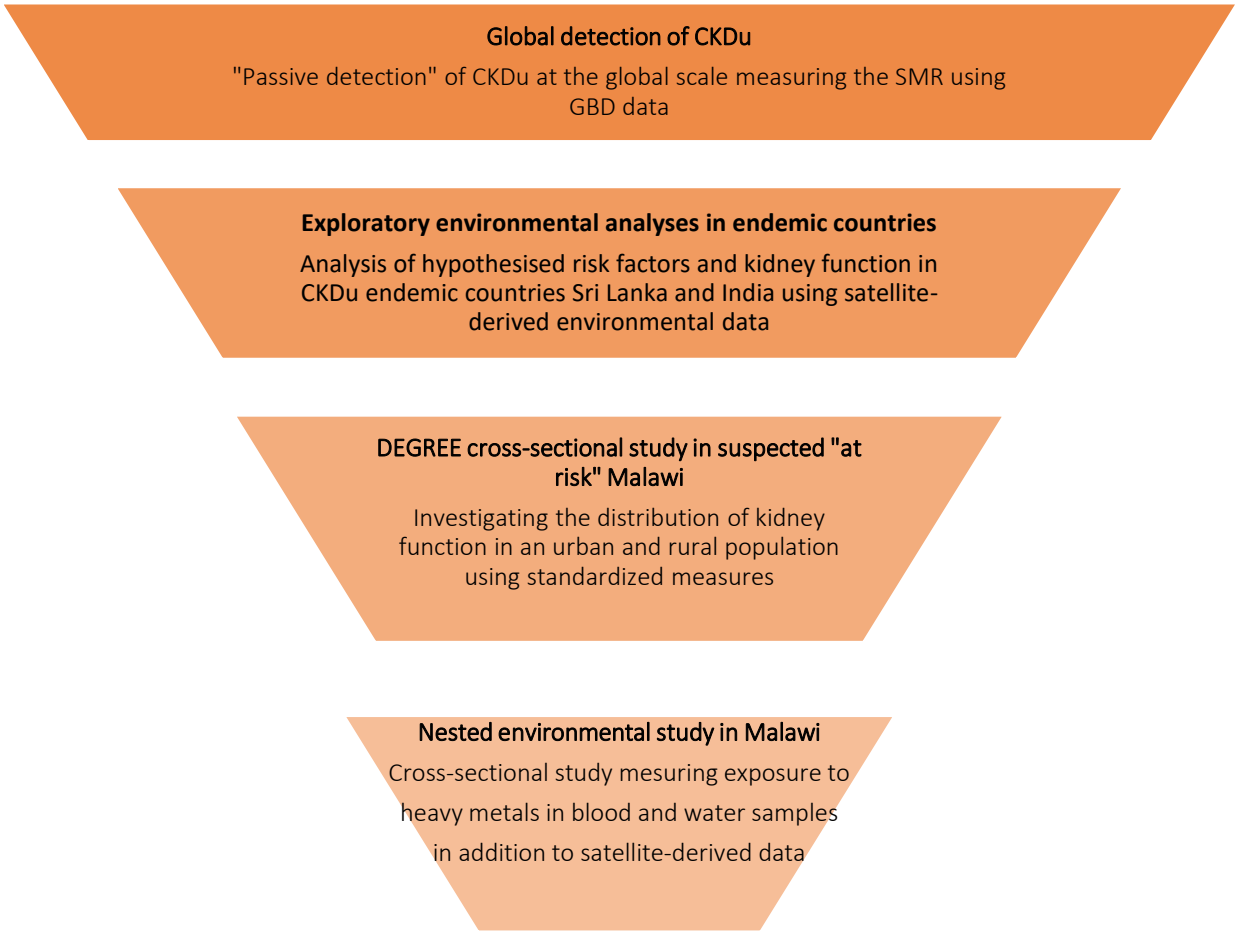


Figure 1. Diagram of the thesis structure showing the “funnel” approach of investigating CKDu from the global level using the Global Burden of Disease (GBD) data, followed by investigation of CKDu in countries which have reported the disease but lack environmental analyses, followed by a cross sectional and nested study of CKDu in Malawi which is suspected to be at risk of the disease

As there are many facets and stages to my research, I present an overview of the data sources and methods used (processing and statistical analyses) used for each chapter in the table below.

Table 1. Summary of data sources and methods used for each chapter in this thesis

	Global detection of CKDu	Exploratory environmental analyses in endemic countries	DEGREE cross-sectional study in Malawi	Nested environmental study in Malawi
Data sources	Global burden of disease database	**Sri Lanka**: cross sectional data collected using the DEGREE standardized protocol (n= 3,296) **India**: cross-sectional data from studies conducted by Public Health Foundation India, and Madras Diabetes Research Foundation (n= 11,119)	Cross sectional data collected using the DEGREE standardized protocol (n= 821)	Nested DEGREE sample (n= 367)
Outcome	Standardized mortality ratio	eGFR <60	eGFR <90	
Data protection measures	None	No permission to show location of participant housing coordinates in maps	Anonymised data using study identification number only. No permission to show location of participant housing coordinates in maps	
Satellite derived environmental data sources, processing and statistical methods	N/A	**Heat index:** ECMWF Era-5 dataset **Processing method:** Bilinear interpolation/Empirical Bayesian kriging (ArcMap v 10.5.1) **Altitude:** Shuttle Radar Topography Mission **Processing method:** Mosaic (ArcMap v 10.5.1) **Land cover:** European Space Agency Climate Change Initiative Land Cover (ESA-CCI LC) **Processing method:** None	N/A	**Heat index:** ECMWF Era-5 dataset **Processing method:** Bilinear interpolation /Empirical Bayesian kriging (ArcMap v 10.5.1) **Altitude:** Shuttle Radar Topography Mission **Processing method:** Mosaic (ArcMap v 10.5.1) **Land cover:** ESA-CCI LC – S2 Prototype Land Cover 20m Map of Africa **Processing method:** None

		Statistical methods for satellite derived variables: Linear and logistic regression with eGFR and study variables and linear mixed modelling		**Statistical methods for satellite derived variables:** Linear and logistic regression with eGFR and harmonized dataset
Field-based environmental data sources, and statistical methods	N/A	N/A	N/A	**Heavy metals - Potable water:** Principal component analysis Pearson correlation matrix Linear regression with harmonized dataset **Heavy metals - Whole blood samples** Bayesian kernel machine regression Linear regression with harmonized dataset

The first part of the thesis is a review of the literature (chapter 2). I will start by describing CKDu prevalence and key environmental risk factors, the issues these studies have presented, and the gaps in the current research field via a narrative literature review. In chapter 3 I will assess global standardized mortality trends of CKDu using deconstructed mortality codes from the Global burden of disease data (GBD). I assess whether the use of such data sources is appropriate in informing global patterns of CKDu prevalence, and whether resource constrained nations should use these types of data to detect signals of disease.

In chapter 4, I investigate CKDu in Sri Lanka and India which have detected CKDu using the DEGREE protocol method, and national cohort data respectively, but have not conducted exploratory individual-level environmental analyses on these populations. In this chapter I conduct an exploratory environmental analysis on the at-risk populations in Sri Lanka and India by assigning temperature, altitude, and land cover variables as a proxy for vicinity to farming practices using satellite-derived data sources.

In chapter 5 I present the results from a cross-sectional prevalence study which was conducted in Malawi in 2018 using the DEGREE protocol methodology to measure kidney function across an urban and rural population. In chapter 6 I present the results from the nested cross-sectional study I conducted in Malawi in 2018 to investigate the association of environmental risk factors with eGFR in a sub-sample of the overarching DEGREE population described in chapter 5. I designed and conducted this nested study, which involved designing questionnaires, sampling frames, data collection techniques, training field workers, water source collection, and supervision of field work whilst in Malawi. I analysed heavy metals in potable water samples which I collected, heavy metals biomarkers in blood, along with questionnaire data and satellite-derived variables temperature, altitude and residential vicinity to and cover in association with low kidney function.

In the discussion (chapter 7) I reflect on my research and put it into the context of the current understanding of CKDu in association with widely hypothesised environmental risk factors. The aim is to discuss how my research has contributed to the growing body of research into CKDu, specifically to aid international comparisons of endemic regions in relation to environmental exposures. I will also discuss how the study I have designed and the

methods I have used can be beneficial to in the wider CKDu research framework. I will then present my remaining research questions and future research recommendations.

Ethics and funding

For this research ethical approval was granted by the National Health Sciences Research Committee Protocol in Malawi, and the London School of Hygiene and Tropical Medicine Ethics Committee protocol. This study was funded by the Medical Research Council Global Challenges Research Fund Foundation Grant (MRC GCRF).

2. Background

2.1 Structure and function of the kidney

The kidneys are paired organs located in the upper abdomen on either side of the spine [38]. Renal tissue consists of the renal cortex and medulla and is surrounded by the renal capsule (figure 2). Oxygenated blood enters the kidneys via the renal artery, and deoxygenated blood leaves via the renal vein. The primary functions of the kidneys are to produce urine via the filtration of waste products in the plasma, and simultaneously control fluid balance, maintain serum concentration of electrolytes and regulate blood pressure [39, 40]. The nephron is the functional unit of the kidney which comprises the glomerulus and tubular system which ends at the collecting duct. The average human kidney contains approximately 1.2 million nephrons[41].

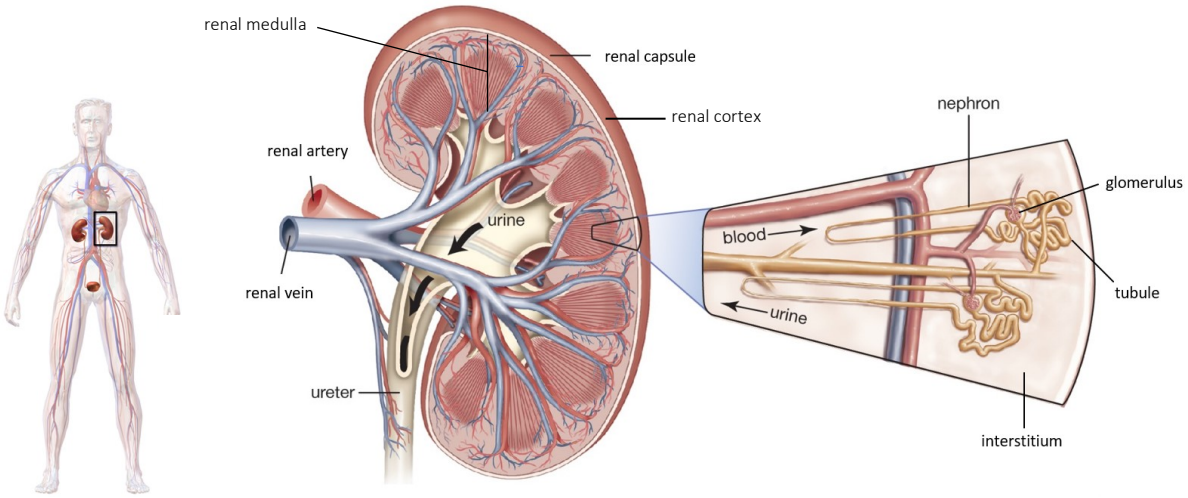


Figure 2. Simplified structure of the kidney (image reproduced from Encyclopaedia Britannica)

The glomerulus is the filtering unit of the kidney, comprising a bundle of looped capillaries lined with fenestrated endothelial cells on the inner surface, and podocytes on the outer surface. The glomerulus is situated between two resistance vessels and is contained within the Bowman's Capsule by a mesh of proteins [42] (figure 3). Blood enters the glomerulus via the afferent arteriole and is filtered through the endothelium and podocytes which are highly fenestrated to allow the passage of small molecules such as water, sodium, and glucose. An ultrafiltrate is produced and collected by the Bowman's Capsule, and the remaining blood

leaves the glomerulus via the efferent arteriole [43]. The capillary filtration barrier acts to retain large proteins and blood cells inside the lumen and separate from the urinary space [44]. Presence of proteins and/or blood cells in the filtrate is usually indicative of a breakdown of this complex barrier, and indicative of renal function loss [44]. Over 24 hours, an average of 180 litres of filtrate is produced and modified through the tubular system to be transformed into final urine. Most of this filtrate is reabsorbed in the tubules, and the average urine output is approximately 1.5-2L per day [45].

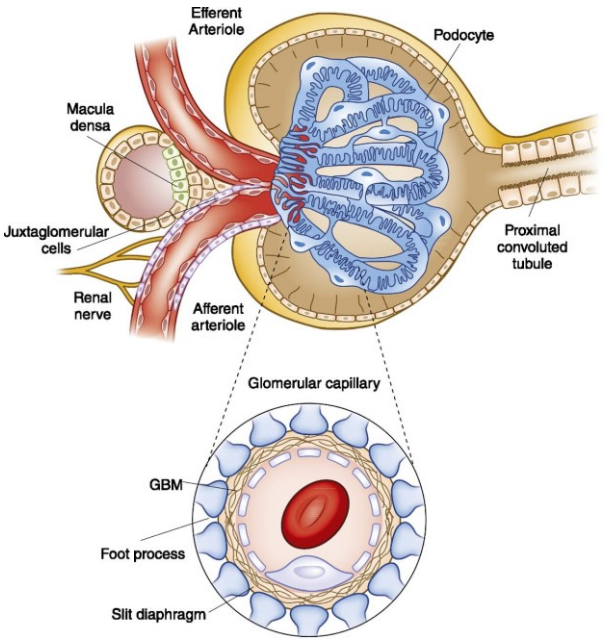


Figure 3. Structure of the glomerulus and Bowman's Capsule

2.1.1.1 Estimated glomerular filtration rate (GFR)

Glomerular filtration rate (eGFR) is a measure of kidney function derived from the blood biomarker serum creatinine (SCr), and is an overall estimation of the volume of blood filtered through the glomeruli per minute [46]. A normal eGFR is classified as >90 ml/min/1.73m^2 body surface area. This value generally declines over the life course due to progressive structural and functional deterioration of the kidney, and is thought to decline by 8ml/min/1.73m^2 per decade [47].

The most common equations used to measure eGFR are the Cockcroft-Gault, the Modification of Diet in Renal Disease (MDRD) and the Chronic Kidney Disease Epidemiology Collaboration

(CKD-EPI) equation; all of which use serum creatinine biomarkers to measure kidney function [46, 48-50]. Creatinine-based estimates however are thought to be influenced by many sources of variation including diet, ethnicity, and body mass index (BMI), potentially leading to overestimation of CKD in a population [49, 51]. As a result, the Kidney Disease Improving Global Outcomes (KDIGO) recommended the use of an additional marker-based method for eGFR calculation which include cystatin C [50, 52] The CKD-EPI equation is the most recently developed and, it is thought to be the most precise when using a combination of cystatin C and creatinine, if compared to creatinine-based estimates alone [46, 53]. It must be noted however, that this equation has not yet been validated for Latin American or Sub Saharan African populations.

2.1.2 Chronic Kidney Disease (CKD)

Chronic kidney disease (CKD) is defined by a decreased eGFR of <60 ml/min/1.73m^2 for at least three months [50, 54]. This case definition categorizes CKD into five stages (G1–G5) and two substages (G3a and G3b) based on eGFR (figure 4). In addition, albuminuria (proteins in the urine) categories (columns A1–A3) are combined with eGFR to create risk profiles (low risk to high risk). Diagnostic investigations ascertain the severity and stage of the disease, with measures of eGFR and Albumin:Creatinine Ratio (ACR), most commonly conducted as preliminary tests [54].

Prognosis of CKD by GFR and albuminuria categories: KDIGO 2012				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 330 mg/mmol	> 300 mg/g > 30 mg/mmol
GFR categories (ml/min/1.73 m²) Description and range	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			

Figure 4. Classification of CKD prognosis by eGFR categories and albuminuria levels as classified by the Kidney Disease Improving Global Outcomes (KDIGO). Risk colour spectrum shows low risk (green); moderately increased risk (yellow); high risk (orange) and very high risk (red) categories

CKD is commonly associated with age-related decline, and co-morbidities hypertension, diabetes and obesity [55], and is characterised by vascular disease, glomerulonephritis, congenital abnormalities, or obstruction of the urinary tract [1].

CKD is described as one of the most neglected chronic diseases [56-58] even though it has consistently ranked in the leading 30 causes of mortality across all socio-demographic groups since 1990 [59]. The Lancet Global Burden of Disease (GBD) Study estimates that the global prevalence of CKD in 2017 was 9·1% which equated to approximately 700 million cases [60]. Historically, CKD has been synonymous with more developed, ageing populations, however regional variations in CKD mortality have been observed, with increasing numbers of deaths now occurring in developing nations [58]. For example, in Central and Latin America, CKD was, respectively, the second and fifth ranked cause of death in 2017 [59, 60]. In these regions there has been an epidemiological shift in mortality from communicable diseases to non-communicable diseases, with major increases in the prevalence of diabetes and hypertension. The GBD data show that the global burden of CKD is increasingly carried by low- and middle-income countries [60]. In some regions, significant changes in disease epidemiology have become apparent; for example, in some Latin American and Asian countries, a newly discovered CKD is now highly prevalent.

2.1.3 Chronic kidney disease of undetermined cause (CKDu)

Over the past two decades, countries across tropic and sub-tropic regions (notably Central America and Asia) have experienced an increase in CKD which is not associated with common CKD risk factors, such as diabetes, hypertension, obesity or glomerulonephritis [1, 4, 5, 14, 61]. This type of CKD mainly affects young, male agricultural labourers conducting high intensity work in high temperatures. Clinical studies have shown that the disease is characterized by asymptomatic increases in serum creatinine, a normal blood pressure, with low-grade proteinuria (protein in the urine). Renal histopathology shows substantial tubulointerstitial and glomerular damage [16, 28, 62, 63]. This chronic impairment of kidney function has been termed CKD of undetermined cause (CKDu). Other terms including Mesoamerican nephropathy (MeN),

Udhanam Endemic Nephropathy, and Rural Endemic Nephropathy have been used when describing the condition across some reporting regions [33, 62, 64].

2.1.4 Challenges of characterizing and diagnosing CKDu

International comparisons have provided the first steps in understanding many chronic diseases[65], but these comparisons rely on the availability of standardised tools to estimate disease prevalence [2]. This is a central issue with CKDu, as the disease is clinically asymptomatic until the late stages, and biases in the methods used to estimate the eGFR in population studies are highly variable across populations. CKDu case definitions across all studies in the literature use the same eGFR threshold to define CKDu: <60ml/min/1.73m^2 in the absence of diabetes, hypertension, and proteinuria [1, 2, 8]. However, quality issues emerge through a lack of information on disease coding in registry data using the International statistical classification of diseases and related health problems (ICD) system, and a disparity between eGFR equations used by those taking bio samples, which could lead to underestimation or misclassification of low eGFR across the study participants.

A clinical definition of CKDu is based on two eGFR measures 3 months apart [54, 66]. This cannot be done in most epidemiological studies, and therefore researchers investigating CKDu tend to measure the eGFR distribution across a study population, rather than diagnosing CKDu itself. Furthermore, as discussed above, the CKD-EPI equation used to calculate kidney function has not been validated in many populations, and therefore needs to be developed for other ethnicities to avoid most frequently overestimation of eGFR in these populations [49, 50]. Historically, CKDu has been diagnosed via a process of elimination. For example, a person living or working in an agricultural community with no known history of kidney disease, diabetes or hypertension but presenting with biomarkers of kidney damage, would be considered a possible case based on epidemiology [67]. Additional information is then required for diagnosis to a probable case of CKDu, such as progression from early non-symptomatic stages to exhibiting symptoms of cramping, dysuria (painful urination), foamy urine, and mild cardiac disorders. A confirmed case fulfils the epidemiological and clinical criteria above, together with a kidney biopsy which demonstrates the histopathological pattern of chronic tubulo-interstitial nephritis. Furthermore, due to CKDu research being in its infancy, there is no clear consensus on the geographical or age distribution of CKDu. It is generally hypothesised that CKDu disproportionately affects males under 50 years living and working in agricultural tropical and

sub-tropical climates. It could be argued that to an extent this geographical characterization created the "streetlight effect" [68] in CKDu research. Such study locations are countries which have a hot climate with reliance on agricultural output. From these issues it is thus clear that there is no standardized method of measuring or defining CKDu.

2.1.5 Prevalence of CKDu

Given these issues with measuring eGFR, the reported prevalence of CKDu is highly variable and depends largely on the methods used in a specific region. What is gleaned from epidemiological studies to date, however, is that CKDu predominantly affects rural populations, and disproportionately, young males [69, 70]. In Central America the prevalence of CKDu in rural versus urban areas ranges from 13% to 26%, and 0% to 9% respectively [4, 14, 71]. A recent study in India reported a significant difference in CKDu prevalence between rural (3%) and urban areas (1%) [5], and a further study in Sri Lanka in an entirely rural population reported an overall prevalence of 6% [6].

In these regions, a significant difference in sex-specific prevalence of CKDu has been observed, with males more frequently affected. In Central American countries, the highest male prevalence estimates range from 13.8%-42%, compared to females which range from 5.8%-22% [4, 71, 72]. In India and Sri Lanka, the CKDu prevalence was markedly higher for males (2% and 11% respectively) than for females (1% and 3.6% respectively) [5, 6]. Generally, agricultural workers are the most affected, specifically those which involved in high intensity labour. Sugarcane workers in Central America have the highest prevalence of CKDu (13-44% for males), and there is evidence from other regions that agricultural labourers are adversely affected[5, 6, 64, 73]. More recently, markers of kidney damage have been found in adolescents that live in the same communities as those with CKDu, but do not conduct agricultural labour [74]. This could suggest that factors other than occupational risk factors for CKDu contribute to the disease – specifically exogenous environmental risk factors.

2.1.6 Leading hypotheses

Many hypotheses have been proposed to explain the epidemiological, clinical, and histopathological pattern of CKDu, however, no clear causal mechanisms have been identified. Vigorous debate between clinical and academic researchers continues about which is the most plausible hypothesis. The leading hypotheses are occupational heat stress and recurrent dehydration [14, 16, 75, 76], pesticide exposure [10, 26, 77], and heavy metals exposure via potable

water[24, 78, 79]. Other hypotheses include household vicinity to cropland (and therefore potential exposure to pesticides) [10] and residing at low altitudes [18, 19].

2.1.6.1 Heat stress

Increasing global mean temperatures have been linked with higher frequency and severity of heat waves which can have a direct effect on human health. One of the consequences of climate-related extreme heat exposure is dehydration and kidney volume loss, leading to acute mortality from exacerbations of pre-existing chronic disease, as well as from direct heat stroke [31]. Recent studies have shown that recurrent heat exposure together with physical exertion and inadequate hydration can lead to CKD in the absence of diabetes, hypertension, or glomerulonephritis. Epidemics of CKDu consistent with heat stress nephropathy are increasing globally [31, 80].

The literature identifies agricultural workers as specifically vulnerable. These workers in lowland areas, are suggested to be at higher risk of recurrent dehydration due to manual labouring in extreme heat and humidity [13, 80]. The observation of similar epidemics across the world has led to the suggestion that "heat stress nephropathy" might represent a form of CKDu [69]. To better understand the mechanisms by which heat stress might cause CKD, animal models have been developed in which mice are exposed repetitively to heat and water restriction. One such study conducted by Jimenez *et. al*. [81] found that dehydrated mice with no access to water lost an average 14% body weight, and those with access to water lost 8% at the end of the heating period. One of the consequences of heat stress and dehydration is the development of elevated serum osmolarity, activating mediator systems that cause chronic tubulo-interstitial disease with fibrosis and inflammation, similar to what is observed in renal biopsies of subjects with Mesoamerican nephropathy across Central America [12]. The mechanism has been linked with development of hyper-osmolarity with activation of the vasopressin and aldose reductase or fructokinase pathway [81]. Recurrent dehydration activates aldose reductase in the proximal tubule, converting glucose to fructose that is metabolized by fructokinase in the proximal tubule, leading to the release of oxidants that cause local tubular injury. Those mice with water access, and those with fructokinase-knockout, were protected from developing renal injury. These studies provide a potential explanatory mechanism for CKDu.

Several human studies in agricultural workers have been conducted across Central America to assess pre-and-post work shift kidney function, that showed similar results to mouse-model studies [13, 29, 82]. A Nicaraguan longitudinal study of kidney function over a 2-month harvest period in 29 sugarcane cutters was conducted with a reference group of non-cane harvesters. Researchers found a clear, significant decrease from pre-shift renal function after 9 weeks of work, with a 16% increase in mean serum creatinine, and a 10% decrease in eGFR. Overall, body weight decreased by an average of 0.3kg (-0.8 - 1.3, P<0.01). Additionally, two new cases of reduced eGFR (<60mL/min) were diagnosed in the cane-cutting group.

A larger cross-shift study conducted across 189 cane cutters in El Salvador [13], measured renal function in association with dehydration, temperature, work time and fluid intake, and found a high prevalence of reduced eGFR (14%), marked by a significant increase in serum creatinine (1.07-1.18mg/dL, p<0.001) (pre-and-post shift) urinary creatinine (from 1.1-1.9g/l, p<0.001) and in serum uric acid (0.7mg/dL (-3.6 – 3.9, p<0.001)). Furthermore, a significant association was discovered between temperature and post-shift increase in serum-creatinine (~2% per degree of heat, (p<0.001). Concentrated post-shift urine suggests individuals were experiencing repeated heavy metabolic loads on the kidneys, via renal tubular absorption. This is in line with other histopathological findings of tubular atrophy and interstitial fibrosis in cases of CKDu detected in Central America, where the disease is known as Mesoamerican Nephropathy (MeN) [28]. Mean weight gain was reported during a workday, with gain positively associated with estimated daily liquid intake, concluding that 1 extra litre of fluid increased post-shift body weight by 0.5kg (p=0.008). This corroborates with Jimenez's finding, that access to sufficient water during and after a workday is beneficial in maintaining kidney volume [81].

Labourers in these studies are working in average ambient temperatures of 30^{o}C. The US Occupational Safety & Health Administration (OSHA) recommends rest for more than 50% of working hours when average temperatures exceed 28^{o}C, and 75% of the times when temperatures exceed 30^{o}c to avoid increased body core temperatures [83]. Many workers do not take appropriate breaks, as they earn on a pay by weight of daily crop cultivation basis. Horton [84] conducted an anthropological study on farm labourers in California, and reported that labourers do not take breaks over concern of loss of earnings, and do not sufficiently rehydrate, as toilets can be over 300 metres from the field, thus hindering their productivity. These studies and reports have highlighted the strong need for preventative measures such as

provision of water, rest, and shade [31] which is a simple, yet effective method to help combat acute effects of heat stress and dehydration in field workers.

There have however been studies which contest the heat hypothesis. An ecological descriptive study conducted in El Salvador, used hospital admission records as proxies for CKDu, and spatial autocorrelation methods. Analyses showed that high temperatures did not appear to strongly influence the location of CKDu-proxy clusters, but instead, that proximity to farmland (especially sugarcane and cotton) was a significant risk factor [10].

Whilst the studies discussed above are valuable, there are significant limitations in their designs. Most notably, the study designs are heterogeneous (hindering international comparison), and sample sizes are relatively small [11], with some not including a referential group for longitudinal studies [13]. Furthermore, subjects involved could have altered their behaviours, such as increasing water intake, and rest during a working day, as they were pre-informed about the risk factors being measured. In the case of the ecological study conducted in El Salvador, the authors placed importance on crop type and risk of CKDu, however they did not use remote sensing to create a spatially explicit map of crop locations, and instead used inference to predict farm size and crop location. Furthermore, data from local weather stations were used. This required significant imputation and interpolation due to missing values, resulting in a potentially less reliable spatial distribution across the raster in comparison to using higher resolution climate reanalysis datasets [10, 11]. Although temperature was not significantly spatially associated with CKDu, the authors acknowledged that heat exposure could still be a likely cause, but not in isolation, thus supporting the "heat-plus" hypothesis, that other external risk factors are exacerbating heat stress-related CKDu.

2.1.6.2 Exposure to pesticides

It has been suggested that the first reports of CKDu in the 1990's coincided with a more extensive exploitation of land for agriculture [85, 86] together with government-subsidised pesticide programmes, to serve exploding international markets [87]. Toxic agents [28, 88, 89] in air, soil and water may be subject to transformation by climate, topography, and soil use. It is hypothesised that these agents can spread to wider farming communities via displacement or absorption into the food chain and can adversely affect susceptible individuals, particularly those with unhealthy lifestyles and harsh working conditions, including dehydration and heat stress [22, 90]. On a molecular level, toxins can affect the kidney tissue through direct tubular cell

toxicity or alterations in intrarenal blood flow, producing secondary tubular damage leading to kidney damage [89, 91]. Based on experimental studies, some common agricultural pesticides including Glyphosate, 2,4-D and Paraquat [92-94] are identified as nephrotoxic however these studies have varying degrees of quality.

Measuring the effects of pesticides is complex. The term "pesticides" is of questionable value as an exposure measure, as this term encompasses thousands of toxins with distinct toxicological actions [95], which may lead to false negative studies [96]. A recent epidemiological review assessing pesticide exposures and CKDu [95] concluded that studies provide scarce evidence for associations between pesticides and regional CKDu epidemics due to poor exposure assessments and study designs. In general, the heterogeneity of study designs, exposure assessment methods, together with bias including selection and recall biases, severely limit the interpretation of positive and negative results between individual studies. Many studies were classified "low" quality for validity due to exposure misclassifications, no adjustment for confounding factors, and potential bias by recruitment of hospital in-patients. Pesticide exposure remains a contentious factor, for these reasons. Furthermore, the lack of causal mechanisms offered by the research serves to weaken the case that pesticide exposure alone is a causative agent. However, concerns in endemic communities remains, and pesticides remain a major subject of debate between researchers.

2.1.6.3 Water contamination

Drinking water is often postulated as a risk for CKDu, with ground water sources being of specific interest due to concerns over high heavy metals concentrations [30, 78]. Metals postulated as etiological factors include arsenic (As), cadmium (Cd), fluoride (F^-), lead (Pb), and hard water ($CaCO_3$) [30, 78, 97-99].

Arsenic is a highly toxic metalloid which accumulates naturally in aquifers through anthropogenic activities [100]. Arsenic is toxic to most biological organ systems, and the most sensitive target organ is the kidney [101], where acute exposure to Arsenic (>10µg/L) can cause tubular necrosis and renal failure [102-105]. Epidemiological studies assessing As exposure at low [106] and high [107, 108] concentrations, suggest a positive dose-response relationship between Arsenic and CKD. In animal studies, mice exposed to As develop glomerular sclerosis, increased oxidative stress and DNA oxidative damage in kidney tissue [109, 110].

Cadmium is extremely toxic and is found in drinking water supplies as a result of industrial waste, or surface contamination by fertilizers. Epidemiological and animal studies have found that the kidney is the main organ affected by chronic exposure, as it stores approximately 50% of the accumulated dose [111-113] causing a decline in eGFR [114-117]. Most studies have investigated the effects of high-level Cadmium exposure (>3µg/L), however, there is some evidence to suggest that low-level exposure may also have a significant negative impact on renal function[118, 119].

The kidney is postulated as one of the primary sites of Lead accumulation [120]. Lead is known to be nephrotoxic at high concentrations (>10 µg/L), but nephrotoxicity at low-level lead exposure has been investigated more recently in epidemiological and animal studies [121, 122]. Acute exposure is known to cause generalized renal tubule defects, whilst chronic exposure leads to tubulo-interstitial nephritis and tubular atrophy [100, 123-125].

Hardness of water is mainly due to elevated calcium (Ca) and magnesium (Mg) [126], and is expressed as $CaCO_3$. The degree of hardness is classified as, soft, moderately hard, hard or very hard when the Ca and Mg content is 0–60 mg/L, 61–120 mg/L, 121–180 mg/L and >181 mg/L, respectively [126, 127]. Although not known to pose a risk to human health, research in CKDu endemic regions of Sri Lanka reports a synergy between hardness and Arsenic toxicity in drinking water, causing renal disease, specifically amongst those who drink water from shallow wells [21, 128, 129].

The majority of the studies in the literature which investigate water contamination in relation to CKDu are conducted in Sri Lanka, specifically in the agricultural communities in the North Central Province (NCP) [73, 90]; some of which suggest that using shallow wells near agricultural irrigation systems could be associated with CKDu prevalence [73]. A study assessing incidence of CKDu and potable chemical water quality was conducted in the North Central Province of Sri Lanka [129] in prevalent and non-prevalent zones. Results showed that in non-endemic zones where residents drink spring water, the concentrations of metals are the lowest. For example, spring water Cadmium concentrations ranged from 0.70-0.55 µg/L, whereas well water had concentrations ranging from 0.83-2.27µg/L. In areas of spring water consumption (>10 years), no CKDu incidence was reported. It is suggested that the low concentration of heavy metal in spring water could have a protective effect against low eGFR. The lack of biomarker data in this

North Central Province study is a limitation, as in this case, making a causal link of water concentrations with physiological effects not possible to measure.

A study which challenges the water hypothesis [130] was a cross-sectional analysis of metals (As, Cd, Pb, and U) in water sources and urine (n=109) in the NCP in endemic and non-endemic communities. Results showed that metals in all sources were below permissible limits [127]; however the metals concentrations in individual urine samples were higher than, and uncorrelated with, those measured in drinking water, suggesting potential exposure from other sources. Mean urinary concentrations of these elements for CKDu cases, were lower than individuals without CKDu both in endemic and nonendemic areas despite the similar water concentrations. For example, mean concentration of Arsenic in CKDu patients was 29.9±22.1 µg/L, which is lower than the concentration measured in individuals without CKDu in endemic (42.4 ± 43.4 µg/L) and non-endemic (39.9 ± 42.8 µg/L) sites. Rather than this being conclusively linked to lower exposures, this result could be due to the kidneys in CKDu patients being compromised in excreting toxic elements, causing further accumulation in the kidney, exacerbating the disease. These results suggest that CKDu cannot be clearly linked to the presence of contaminants in drinking water alone but could be linked to exposure from other sources.

2.1.6.4 Altitude

Altitude is hypothesised to have protective effect against reduced eGFR [14, 18]. From a physiological perspective, the hormone Erythropoietin (EPO) is thought to help slow renal disease progression, as it attenuates interstitial fibrosis and reduces apoptotic (programmed) cell death [131]. Apoptotic cell death is a contributor to CKD progression, meaning that by reducing cell death, EPO might lead to slowing of the progression of CKD [18, 132]. Furthermore, there is a well-defined relationship between altitude and EPO, which increases in response to low oxygen concentrations at high altitudes. A study conducted in the US using National Health and Nutrition Examination Survey III data, sampled residents living in altitudes from 9-3695 feet (stratified into 5 categories), and analysed altitude in relation to eGFR, controlling for potential confounders including age, sex, income, education and urban versus rural residence [18]. Results revealed that mean eGFR was significantly higher at altitudes >200 feet, when compared with lower altitude categories (78.4 ± 21.6, 85.4 ± 26.8 feet for categories 1 and 5, respectively; $P < 0.05$). Additionally, relative risk of eGFR <60 was lower at altitudes >200 feet

(0.94 versus 1, for altitude category 3 versus category 1, $P < 0.05$)[18]. Limitations include no consideration of duration of residence, land cover at differing altitudes, or occupation at a specific altitude. Furthermore, temperature was not considered as a risk factor in this study., which is a limitation as it is an important variable when assessing altitude together with heat stress (a major causative hypothesis of CKDu). It is however plausible to suggest that altitude could be a surrogate for temperature, and communities at higher altitudes could be experiencing lower prevalence due to working in lower ambient temperatures and humidity levels.

As many believe CKDu is an occupational disease, it is important to incorporate occupation type, into analyses. One such study was conducted by Peraza via a cross-sectional population survey in El Salvador, examining serum creatinine and eGFR in 5 communities at differing altitude and demographics [14]. Results showed a gradient for elevated serum creatinine levels with years of coastal sugarcane or cotton work in males and females, with the highest prevalence found in coastal, sugarcane communities (males 28%, 32%, women 16%, 13%). Sugarcane and coffee communities at higher altitudes had a lower prevalence of elevated serum creatinine (males 4% and 0%, females 5% and 2%). Urban prevalence was intermediate (men, 11% women 8%). In coastal communities, the prevalence for men aged 40-49 (46%) neared the prevalence for older men aged 50-60 (52%). Furthermore, the prevalence of eGFR <60 was highest in the two coastal sugarcane communities (men 19%, 18%, women 8%, 8%). Proteinuria was low grade and infrequent, which is compatible with CKD originating in the tubulo-interstitium, as opposed to glomerular damage [4]. The interpretation of the above studies requires disentanglement, as it is generally unclear whether the higher altitude itself exerts the protective effect against low eGFR due to lower humidity, or whether there is a difference in land cover percentage between lower and higher altitudes which could be a proxy for lower intensity farming practices. A valuable further investigation into this relationship would be to create spatially explicit crop maps at different altitudes, to compare land cover class and cover

There is however a biological counter argument to the hypothesis that increasing altitude has a protective effect on eGFR. Some studies have shown that living at high altitudes is risk factor for kidney damage as it can cause high altitude renal syndrome (HARS), which results from chronic hypoxia that is characterized by polycythaemia, hyperuricemia, hypertension, and

proteinuria. A cross-sectional study conducted in Costa Rica which investigated CKD, found a negative association and increased risk between increasing altitude and low eGFR [19]. However, this syndrome has generally been described at altitudes higher than 2400 meters, and is linked to hypertension and proteinuria, which would suggest that it is not related to CKDu.

A final aspect which is not studied in my thesis, but which may impact one's susceptibility to kidney disease is the role of genetics. This area of research is gaining traction, with an increasing number of studies being conducted on genome sequencing and investigative cohort studies [133, 134]. A recent case control study (n=8 + 8) conducted in the North Central Province of Sri Lanka investigated the involvement of genetic factors in CKDu pathogenesis by conducting whole-exome sequencing in to elucidate the genetic variants associated with CKDu. This association study reported a genetic variant in the KCNA10 gene (rs34970857) which encodes a voltage-gated potassium channel were found in CKDu cases. Furthermore, four variants were identified in gene encoding Laminin beta2 (LAMB2) which is known to cause congenital nephrotic syndrome. Other studies however have reported that no single gene is responsible for genetic susceptibility of CKDu, however this disease could be a result of polymorphism with wide allele frequencies[91, 135].
These results suggest that genetic investigations provide promising and interesting evidence of the presence of genetic susceptibility for CKDu which will require further, targeted research in future studies, which could also be incorporated in tandem with environmental exposure studies.

Table 2. Summary of key papers used in the literature review

Authors	Design	Country/region	Sample size/sex			Exposures	Outcome definition	Findings	Quality comments
			n	Male	Female				
Jimenez, et al 2014	Dehydration protocol in Wild-type and fructokinase-deficient mice	United States, Colorado	36	36	0	Heat	Decline in eGFR and body weight	Dehydrated mice with no access to water lost an average 14% body weight, mice with access to water lost 8% at the end of the heating period	High quality - Robust functional measures (including eGFR, body weight) - Elucidation of dehydration mechanisms which can translate into future human cohort studies
Wesseling et al. 2016	Cross-shift longitudinal	Nicaragua, León and Chinandega	29	29	0	Heat	Decline in eGFR over a harvest period	Significant decrease kidney function, after harvest. Cross-shift increase in serum creatinine may be caused by chronic dehydration. Referent group: no increase in uric acid, weight gain observed	Low quality - Small sample size affects statistical power of results - Different follow-up timing between groups - Blood and urine samples not collected from cane cutters at immediately post-shift potentially allowing workers to rehydrate - Analyses performed on frozen samples long-term storage can decrease concentrations of biomarkers including NGAL
			Referent sample 25	25	0				
VanDervort et al. 2014	Descriptive ecological	El Salvador, National	24,726 (sex not detailed)			Heat, vicinity to cropland	eGFR<60	High temperatures did not appear to strongly influence the location of CKDu-proxy clusters, but instead, that proximity to farmland (especially sugarcane	Medium quality - No biomarker data used, instead use of routinely collected data for CKDu hospital admissions - Issue of misclassification of CKDu - Local weather station data used which does not have uniform distribution of points therefore interpolation could have

						and cotton) was a significant risk factor	higher prediction errors and affect regression results with CKDu admissions data
Laws et al. 2016	Cross-shift longitudinal in 7 occupational categories (agricultural and non-agricultural)	Nicaragua, North West	284 251 33	Occupation	Decrease in kidney function over harvest	Urinary levels of NGAL and IL-18 increased during harvest season. Cane cutters are at highest risk for kidney injury during the harvest.	Medium quality - Use of biomarkers is a strength - Job category used as a surrogate exposure metric which represent one or more unidentified risk factors - which may or may not be related to occupation itself - Loss to follow up during harvest season
Peraza et al. 2012	Cross-sectional	El Salvador	664 256 408	Occupation	eGFR <60	In coastal communities, 30% had elevated serum creatinine, and 18% had eGFR <60 mL/min/1.73 m2 compared with 4% and 1%, respectively, in the communities above 500 m. Prevalence of elevated serum creatinine and eGFR <60 were highest for coastal sugarcane and cotton plantation workers	Medium quality - Descriptive, cross sectional design meant that participants could not be diagnosed with CKD/u thus bias could be introduced into results
Jayasumana et al. 2015	Case control	Sri Lanka, North Central Province	Cases 125 89 36 Controls 180 98 82	Occupation, heavy metals in potable water, application of pesticides	eGFR <60	Higher risk of CKDu for farmers, well water drinkers, history of drinking from abandoned wells and spraying glyphosate. Significantly higher amount of hardness, glyphosate, and heavy	Low quality - Recall bias could affect results of the participant questionnaire on lifestyle risk factors - Controls were not based in the same community as cases, and were not matched - Level of pesticide exposure was not measured, nor were details application methods and use of PPE

						metals levels in abandoned wells. Reservoirs in the endemic area showed contamination with glyphosate but at lower levels.	
Rango et al. 2015	Cross sectional	Sri Lanka, North Central Province	134 82 52	Heavy metals in potable water	eGFR <60	Levels of inorganic nephrotoxicants in all water sources were below permissible limits for drinking water. No significant difference observed in concentrations of metals across drinking water sources from endemic and non-endemic zones. Inorganic contamination of drinking water sources unlikely to be the cause of CKDu here	Medium quality - Cross-sectional data collection in a small sample and changes longitudinal drinking water habits may introduce bias into results. - Participants not randomly selected and were at advanced stages of the disease and are likely not representative of CKDu patients in endemic areas.
Wasana et al. 2016	Random sampling of water sources across high-low prevalence CKDu regions	Sri Lanka, North Central Province	Water sample n = 564	Heavy metals in potable water	eGFR <60	In non-endemic zones where residents drink spring water, the concentrations of metals are the lowest. In areas of spring water consumption, no CKDu incidence was reported. It is suggested that the low concentration of heavy metal in spring	Low quality - No biomarker data used, and no confounders considered - Sampling conducted between wet and dry season – not possible to assess changes in water quality over time

						water could have a protective effect against low eGFR.	

Many studies are limited by exploring only pre-conceived hypotheses and thus lack a multifactorial approach. This issue can be overcome by conducting preliminary descriptive exploratory studies across endemic areas, which can serve as the basis for future research including cohort studies, detailed exposure assessments, and intervention studies. However, ecological studies are commonly approached with caution due to the "ecological fallacy", [136], whereby factors associated with national disease rates may not be associated with the disease in individuals [137]. Ecological studies were considered a remnant of pre-modern epidemiology, however, it has re-emerged, as a plausible, and informative research methodology with potential long-term benefits [138]. Population-level studies are instrumental in defining public health issues, and generating hypotheses for their potential causes [138].

Diez-Roux suggested that failure to account for the community-level context as an effect modifier and determinant of individual level exposures (termed "individualistic fallacy") [139] results in undue attention being focussed on individual characteristics. Multifactorial investigations are effective and preferable, as due to the ability to analyse large proportions of a population they allow for estimation of broader population-level effects, whilst also including individual level effects – thus avoiding ecological and individualistic fallacies [138, 139].

2.1.7 The need for standardization – the DEGREE study

Having presented the main facets of CKDu and the main challenges faced in (i) detecting the disease and (ii) investigating potential environmental risk factors at the individual level, it is necessary to establish a defined and agreed set of criteria for public health surveillance in different regions of the world with the same epidemic pattern. Common issues raised in the literature relate to study design heterogeneity, and the need for measurement standardization [74, 75, 140]. It is imperative to obtain verified, reliable eGFR measures before global comparative analyses can be made. This will allow comparative standardised reporting from differing geographic locations with similar disease patterns [85]. The DEGREE survey is an international protocol created by clinicians and epidemiologists at the London School of Hygiene and Tropical Medicine [1]. This survey aims to estimate the distribution of glomerular filtration (GFR) in an urban and rural population via a cross-sectional prevalence and risk factor study. This involves the quantification of renal function in a representative adult population-based sample and a requirement for standardization of serum creatinine measurements, along with sample

storage for cross centre validation of serum creatinine values. This overarching study will be discussed in more detail in chapter 5. A major factor missing in the scope of this study however is detailed analyses on environmental risk factors in suspected endemic regions. As discussed throughout this chapter, there has been an inherent lack of thorough investigation into exogenous risk factors for CKDu at the individual level. Environmental studies to date have had heterogenous study designs, and a majority have not included bio samples in their analyses, thus making any causal inference a challenge.

2.1.8 Summary

In summary, little is known about the potential aetiologies of CKDu, however several possible risk factors have been suggested. There is a clear need for prevalence studies to be conducted using standardized measurement methods to identify high-risk areas. Furthermore, as there is currently no indisputable evidence to identify possible environmental causative factors for CKDu, there is also a need to develop more in-depth environmental exposure studies in identified high-risk areas. The studies I have conducted and present in my thesis take a step in addressing this gap in the current CKDu knowledge.

3. Passive detection of CKDu using the Global Burden of Disease data

3.1 Introduction

Cause-specific mortality is a fundamental metric of population health. Trends in causes of death can provide important estimates of the burden of disease from the global to national level. [59, 141, 142]. Chronic kidney disease (CKD) is described as one of the most neglected chronic diseases [56-58] even though it has consistently ranked in the leading 30 causes of mortality across all Socio-demographic Index (SDI) groups since 1990 [59]. The net health burden of CKD is complex and shaped by upstream environmental and socioeconomic disparities which, coupled with limited epidemiological data makes detailed quantitative analyses of the disease particularly challenging [58, 59].

In the case of CKDu, epidemiological data is even more scarce, as most accounts of the disease are anecdotal and based on individual reports. Very little is therefore known about its epidemiology, particularly whether it is a global epidemic, or a collection of regional diseases exacerbated by local aetiologies. As discussed in chapters 1 and 2, CKDu has been reported predominantly across Central America (El Salvador and Nicaragua) and South Asia (Sri Lanka and Southern India) and disproportionately affects male agricultural workers [6, 14, 30]. However, there is a lack of consensus on the case definition(s) for CKDu making a comparison of the disease across populations a challenge. Currently, there is a need [143] to discover which parts of the world are experiencing endemic and /or epidemic CKDu (as per current diagnostic criteria) by applying multiple layers of analyses from the global to the individual level to identify at-risk regions and populations. A first step is to measure disease mortality at varying geographical levels (known as passive detection) which can be used to inform future surveillance at the individual level (active detection) [2].

The International Society of Nephrology's International Consortium of Collaborators on Chronic Kidney Disease of Unknown Etiology (i3C) [2] have created a working group providing a common approach to the passive and active detection of CKDu [15]. The i3C suggest a passive detection approach based on analysis of routine data sources, or "minimum datasets" [2] to estimate the extent of CKD in a region of interest and determine the proportion of that estimate that is attributable to non-diabetes (DM) related CKD as a proxy for CKDu. With these ecological data, it is hoped that hot spots of CKD which are not attributable to diabetes can be

identified and targeted for active detection methods [2]. This approach could be a useful first step in resource-poor areas where population surveys may be too costly to conduct without prior ecological analyses and collaboration with local clinicians.

Historically, geographical comparisons of mortality data have been a key component in theory and hypothesis generation. For example, the international comparison of cancers [144-146] in the 1950s and 1960s provided epidemiological information which is still used by current researchers to generate hypotheses and biological discoveries. It is argued that without these initial population-level analyses, the more recent discoveries from these diseases might never have been investigated [138]. These types of analyses should however be approached with caution as they are descriptive and cannot usually be used to assess specific risk factors due to the "ecological fallacy"[137], coupled with ascertainment and misclassification bias in datasets. When classifying renal mortality, the primary cause of death is usually the result of co-morbidities such as infection from Haemodialysis, or complications from diabetes, making the use of mortality data to measure the burden of CKD more difficult. For CKDu this is even more complex as the disease is unlikely to be recorded on death certificates due to the lack of diagnostic consensus.

The Global Burden of Diseases, Injuries and Risk Factors Study (GBD) is a major effort to collect and compile data from vital registration systems, hospital records and verbal autopsies. By applying complex statistical modelling [59] to these data there are now estimates of the burden of CKD which identify secular trends and disparities between countries. These data have enabled epidemiologists, clinicians and policy makers to direct healthcare and interventions more effectively [58, 147]. Despite this major advancement however, the global assessment of renal mortality remains an analytical challenge. This is largely due to variation in certification skills among physicians, the diagnostic pathological data available at the time of completing the death certificate and differing medical cultures in choosing the underlying cause(s) [141, 148]. These issues directly affect the quality of input data to the GBD and can lead to non-representative results.

The topic of mortality data quality is debated across the clinical and epidemiological fields. Studies on verbal autopsy and death certificate quality show a wide global disparity between reporting regions[2, 59], with two major issues arising when using these data for attribution of cause of death analyses. Firstly, there is difficulty differentiating whether recorded kidney

disease is due to CKDu via verbal autopsy and/or death certificates, and misattribution of cause of death is a major challenge worldwide, but specifically in settings where renal biopsies are not routinely conducted. Thus, there is potentially widespread misclassification of CKDu as many disorders of the kidney (both acute and chronic) are not specified by a specific clinical characteristic. Secondly, the attribution of a kidney disease-related death may be non-representative or incorrect as both known causes of CKD and causes of CKDu may co-exist, thus increasing the risk of misclassification during verbal autopsy in such cases[149].

The World Health Organization (WHO) and Institute for Health Metrics and Evaluation have published global estimates on cause-specific deaths including those attributed to kidney disease[150]. Figure 5 is a map of global cause of death data-quality, which includes estimates based on verbal autopsy-based diagnosis of CKD[150]. From this map, it is clear that in almost all regions likely to be impacted by CKDu, data quality is suboptimal.

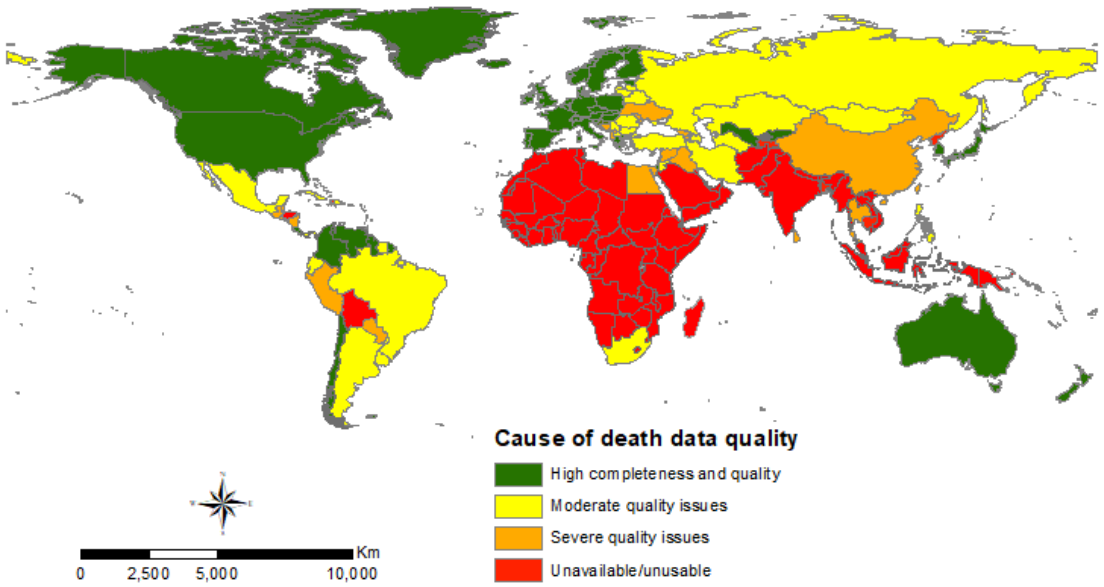


Figure 5. Map showing data quality of cause of death according to the World Health Organization Global Health Estimates Project

Data validation is fundamental to effective clinical and epidemiological research. With the increasing emphasis on research using secondary data sources, validating the accuracy of underlying information of individuals with kidney disease is becoming increasingly important. A systematic review has been conducted on approximately 30 studies[151] which aimed to validate the accuracy of underlying data relevant to kidney disease across multiple higher

income countries including the United States, Italy and Israel. From the data, sensitivities and specificities for (i) presence of kidney disease (ii) cause of kidney disease (iii) comorbid conditions with kidney disease, and (iv) cause of death in kidney disease were presented. Most studies included in this review investigated the accuracy of kidney disease reporting, comparing ICD coded kidney disease to kidney disease defined by the estimated glomerular filtration rate (eGFR). Sensitivity of coded renal disease varied widely (0.08–0.83). Specificity was higher, with all studies reporting values of ≥0.90.

There were two U.S. studies which addressed the accuracy of the variable "cause of death" in individuals with end stage renal disease. [152, 153]. Perneger et al.[152] compared death certificate reports to the state renal registry data, reporting a low agreement of 31% between the sources, with a kappa value of 0.24. Furthermore, Rocco et al.[45] compared death classifications from a national prospective randomized controlled trial of prevalent hemodialysis patients with the Health Care Financing Administration database of death records, where agreement varied considerably by cause of death, and no overall agreement was calculated. This systematic review is one of very few which assesses the validity of CKD mortality data and shows that (albeit in higher income nations) routinely collected data sources need scrutiny before use in any research effort. There was a wide range in accuracy of kidney disease-related variables, and additional research is required to investigate the sources of this variation across higher and lower income nations collectively. It was suggested by the authors that initiatives to improve the quality of underlying kidney disease data must be implemented, including standardized reporting for prevalence estimates. Perhaps the most pressing of all is to increase healthcare givers and clinician's recognition and documentation of disease, through more systematic and rigorous testing and use of standardized ICD coding[151].

Issues with data collection for renal failure were highlighted in a recent survey conducted by the International Society of Nephrology (ISN) on country capacity and response for kidney care across 130 countries (many of which inform the GBD database). Public funding structures for all forms of renal replacement therapy (RRT) (haemodialysis, peritoneal dialysis and kidney transplantation) were much less common in low-and lower-middle income regions, and free access to renal care medications were non-existent [154]. In contrast to this, over one third of high-and upper-middle income nations reported complete funding for medications in non-

dialysis kidney care [154]. As a result of these issues, there was limited availability of health information systems including renal and mortality registries, specifically for CKD and acute kidney injury (AKI) across most low and lower-middle income regions. Similar findings have been reported in other global RRT studies, which suggest that availability of treatment for end stage renal disease (ESRD), and thus comprehensive mortality data is focussed on patients who are able to access RRT in high and upper-middle income regions, with a deficit in low and middle income nations [147, 155-157].

Furthermore, RRT is not a robust proxy for CKD prevalence for three reasons: firstly, the majority of patients who develop CKD in low-income countries are never offered RRT, secondly, most of those who do receive RRT are removed from therapy or die, and finally, RRT only relates to end stage renal disease due to severe CKD [158]. Most importantly however, current registration data do not capture the large proportion of patients who die without ever accessing care. An assessment of RRT in developing regions will therefore underestimate the true burden of CKD, and the extrapolation of other renal conditions [159].

Figure 6 shows the global prevalence of treated end stage renal disease per 1 million population in 2016. Prevalence roughly reflects the economic status of a country, with a higher proportion of treatment observed in higher income nations including the USA and parts of central Europe and East Asia.

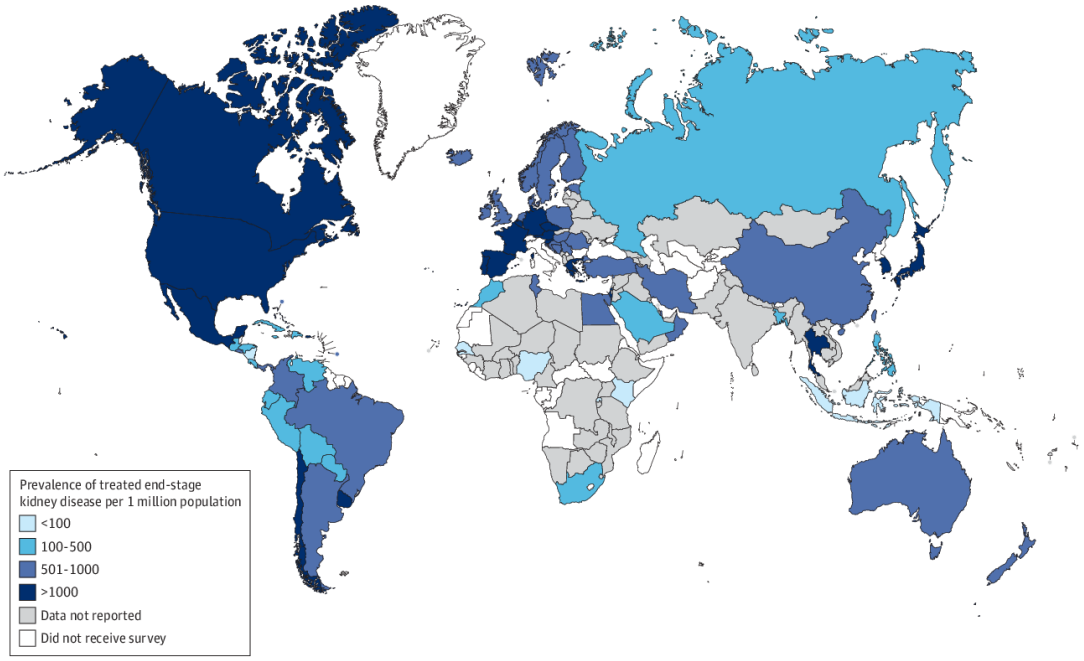


Figure 6. Global prevalence of treated end stage renal failure (ESRD) (per 1^6 population) based on individual country data. Source: Bello et.al, 2017

As discussed above, there are undoubtedly significant limitations in using routinely collected data to estimate CKD and to extrapolate signals of CKDu. However, CKDu research is in its infancy, and currently a multi-level approach must be applied to CKDu data in order to detect signals of the disease starting from the global level (using global CKD mortality data), to the individual-level (using bio sample data from cross-sectional and longitudinal cohort studies).

In this chapter I apply the i3C "passive detection" method to estimate the global extent of CKD mortality using the GBD database. From these estimates I determine the proportion of CKD mortality which is not attributable to DM (a proxy for CKDu) to the overall CKD estimates for 2017. I then calculate the standardized mortality ratio (SMR) for each cause of death and present the results in sex and age-specific disease maps. Finally, in order to detect whether the CKDu signal is higher in proportion to the DM-related CKD, I map the rate ratio of non-diabetic CKD (or CKDu) to DM-related CKD. This will denote whether the rate of non-DM CKD is in fact higher than that of the CKD attributed to diabetes in a country, which would show there is a positive signal for the disease.

I aggregated countries according to World Bank income groups "High income" and "Upper middle income" versus "Low income" and "lower middle income" to reflect the availability of RRT [(2)], and therefore the likely quality of mortality data.

The aim of this geographical analysis is to assess whether routinely collected data is suitable for the detection of CKDu when applied to the i3C passive detection method, and to highlight potential high-risk regions which could serve as a basis for future active detection.

3.2 Methods:

3.2.1 The passive detection method – an overview

The passive detection of CKD death rates is done by modelling age-standardized death rates attributable to CKD from national mortality registries. The numerator is deaths attributable to CKD, and the denominator is the mortality for the population at risk. In order to yield greater specificity of the detection of CKDu, I subtracted the deaths attributable to DM CKD from overall, "all cause CKD" deaths. This is done because DM is the most common cause of CKD

(with or without hypertension) and can help to separate "traditional" CKD from the non-DM related CKDu (or proxy for CKDu) [2].

3.2.2 Data sources

The 2017 GBD Study captured and summarized data from 1990-2017 on death and disability for over 350 diseases and injuries across 195 countries. The study includes analysis of CKD based on the 10th revision of the standardized international classification of diseases (ICD-10) coding of hospitalization and mortality, and impaired kidney function [160]. Detailed information on GBD ICD-10 modelling methodologies have been described elsewhere [59, 142]. In order to calculate an "all-cause CKD" rate, I extracted and combined mortality codes B8.2 ("Chronic kidney disease") and B8.3 ("Acute glomerulonephritis") from the GBD database. After discussion with the nephrology team at the London School of Hygiene and Tropical Medicine, it was decided that acute glomerulonephritis would be included in the "all cause CKD" category as although it is an acute disease, it is often misclassified. Therefore, this acute disease category was included in order to account for misclassification. I also extracted B8.2.2 ("Chronic kidney disease due to diabetes mellitus type 1") and B8.2.3 ("Chronic kidney disease due to diabetes mellitus type 2") in order to calculate the rate of CKD attributed to DM (or "CKD attributed to diabetes"). I extracted mortality codes which were assigned to the 195 GBD countries and territories in 2017. I stratified data by sex and target age categories "15-49 years", "50-69 years" and "70 plus years", which were the available age categories for these codes for 2017. Although CKDu is thought to be a disease disproportionately affecting working-age populations, I wanted to investigate whether the SMR values for CKDu would be higher in the oldest age categories in comparison to the younger categories in order to decipher whether the signals detected for CKDu were actually indicative of CKDu, or likely misclassifications for DM-related mortality. Children under the age of 15 were also not included in these analyses due to the very small number of CKD-related deaths in this age category within the GBD database. Population data were obtained from the GBD 2017 population estimates which are stratified by age and sex and income category (2017) [160]. Table 3 shows the GBD mortality codes used for these analyses.

Table 3. Table showing the 10th revision of the standardized international classification of diseases (ICD-10) codes, and respective Global Burden of Disease re-coding categories used to estimate the SMR values for: "all cause CKD", "CKD attributed to diabetes" and "non-diabetes related CKD"

Mortality measure	ICD-10 CODE	GBD RE-CODE
Chronic kidney disease	D63.1, E10.2, E11.2, I12-I13.9, N02-N08.8, N15.0, N18-N18.9, Q61-Q62.8	B8.2
Chronic kidney disease due to diabetes mellitus type 1	E10.2	B8.2.1
Chronic kidney disease due to diabetes mellitus type 2	E11.2	B8.2.2
Chronic kidney disease due to hypertension	I12-I13.9	B8.2.3
Chronic kidney disease due to glomerulonephritis	N03-N06.9	B8.2.4
Chronic kidney disease due to other causes	N02-N02.9, N07-N08.8, N15.0, Q61-Q62.8	B8.2.5
Acute glomerulonephritis	N00-N01.9	B 8.3
Aggregation of GBD recoded ICD-10 codes for CKD SMR analyses		
"All-cause CKD"		B.82 + B8.3
"CKD attributed to DM"		B8.2.1 + B8.2.2
"Non-DM related CKD"		Sum (B.82 + B8.3) – Sum (B8.2.1 + B8.2.2)

It must be noted that the GBD does not include mortality data for "acute kidney injury", "Chronic tubulo-interstitial nephritis", or "Tubulo-interstitial nephritis, not specified as acute or chronic". These causes of death would be entirely appropriate codes for detecting CKDu, even though they would be in lower numbers. This could therefore result in underestimation of the excess deaths from CKDu.

3.2.3 Statistical analyses

I calculated an "All cause CKD" mortality estimate by merging the counts of codes B8.2 and B8.3, and a "CKD attributed to DM" mortality estimate by merging the sum of B8.2.1 and B8.2.2. I calculated a "Non-DM-related CKD" estimate (CKDu proxy) by subtracting the "CKD attributed to DM estimate from the "All cause CKD" estimates. I then modelled the age and

sex-specific standardized mortality ratio (SMR) for 2017 for each mortality code and country. The SMR is a commonly used comparable statistic which estimates the ratio of observed and expected CKD deaths. The global age and sex-specific rates for each CKD code were used as the standard comparator for each World Bank income group. As mentioned above, there is a disparity between availability of RRT in upper-middle and lower-middle income nations which will likely reflect the quality of data which is fed into the GBD. As CKD death rates are a conflation of CKD burden and RRT availability, there is significant variability in the capacity for renal care across upper- and lower-income countries, making a direct comparison between these groups challenging. Due to this issue, coupled with the disparity between SMR and rate ratio values for each income group, I mapped the CKD mortality outcomes to separate low- and high-income scales to facilitate international comparisons within each income group. I calculated the SMR using the indirect method, as:

Equation 1. SMR formula

$$SMR = \frac{O_j}{E_j} \quad (161)$$

Where:
O_j – Observed deaths in the geographical unit (j) in 2017
E_j – Expected deaths in the geographical unit (j) in 2017
j – Geographical unit (country level)

The formula applied to calculate expected deaths was:

Equation 2. Expected deaths formula

$$E_j = \sum_{i=1}^{k} \frac{D_i}{P_i} * P_{ij} \quad (161)$$

Where:
i – age group (1 to k).
D_i – Global CKD deaths in age group (i) 2017 stratified by income
P_i – Population of age group (i) 2017
P_{ij} – Population of age group (i) in geographical unit (j) 2017

The 95% confidence interval (CI) was estimated based on the expectation that the number of observed events followed a Poisson distribution [162]. The global SMR distribution for each mortality code was represented in age and sex-specific disease maps categorised into World Bank Income groups "High income" and "upper middle income" (figure 7) (now referred to as "high income") versus "Low income" and "lower middle income" (now referred to as "low income") which reflects the availability of RRT in these regions.

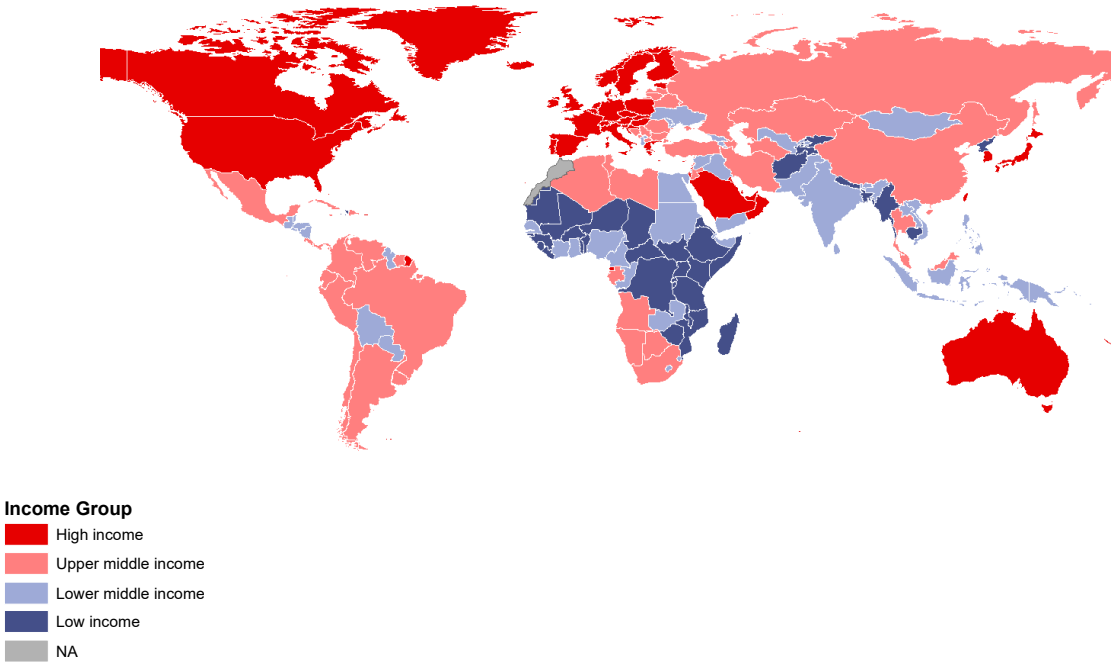


Figure 7. Global distribution of World Bank income groups by country

In order to clarify whether the "non DM related CKD" signal is indeed indicative of true CKDu, I calculated a rate ratio of "non DM related CKD" to "CKD attributed to DM" (Non-DM-CKD:DM-CKD) per 100,000 population. If the rate ratio remained high in the countries which had elevated SMR values for "non-DM related CKD", this would likely suggest that the signal detected is highlighting a country or region which is at high risk of CKDu.

All results were mapped using age, sex and income-specific cartograms.

All statistical analyses and mapping were performed using R Studio version 3.5.0.

3.3 Results

3.3.1 Global spatial patterns of SMR

Figures 8a-19c show age and sex-specific SMR mortality trends for “All cause CKD”, “CKD attributed to DM” and “non-DM-related CKD” for high- and low-income countries in 2017. An SMR value lower than 1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark. Tables show the countries which had the highest SMR values, which have been grouped into regions as per the GBD classification.

3.3.2 High income countries

Males 15-49 years

For all CKD-related mortality outcomes, excess mortality was concentrated in tropical and sub-tropical countries in Southeast Asia, Oceania, Central America and the Caribbean. Table 4 shows the countries which had the highest excess mortality for these regions.

For “all cause CKD”, excess mortality was generally higher across Southeast Asia (Mauritius SMR=4.13 95%CI=3.64-4.62) and Oceania (Marshall Islands SMR=3.02 95%CI1.90-2.70). The values observed in Central America (Mexico SMR=2.56 95%CI=2.10-3.21) and the Caribbean however (Dominica SMR=2.37 95%CI=2.10-2.44) were comparatively lower (Figure 8a).

Excess mortality attributable to “DM-related CKD” was lower than that for “all cause CKD” across all countries in this income, age and sex category. All countries in Central America and the Caribbean had values significantly below the global benchmark, whilst countries across Oceania and Southeast Asia had the highest comparable excess mortality for this category (table 4, figure 8b).

Excess mortality for “no-DM related CKD” was of a higher magnitude than for the other causes of death in this age group. The highest SMR was observed in Mauritius (SMR=12.98 95%CI=10.89-15.06). Furthermore, in Central America and the Caribbean, SMR values for diabetic CKD were very low in comparison to the non-DM CKD mortality, which suggests that the signal in these regions is potentially stronger than the other regions (table 4, figure 8c). These higher values observed for “non DM related CKD” in this sex and age group corroborate with the literature that CKDu occurs in working aged males in tropical and sub-tropical countries and could be indicative of a positive signal detection of the disease[14, 31, 76, 77].

Table 4. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: “all cause CKD”, “CKD attributed to DM” and “non-DM related CKD” for males 15-49 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to DM	Non-DM related CKD
			SMR (95% CI)	
Southeast Asia	Mauritius	4.13 (3.64,4.62)	2.30 (1.90, 2.70)	12.98 (10.89, 15.06)
	Seychelles	2.64 (1.48, 3,22)	1.26 (1.21, 1.38)	9.26 (9.00, 11.41)
	Sri Lanka	1.17 (1.10, 2.34)	0.58 (0.44, 0.76)	4.01 (3.49, 4.56)
Oceania	Marshall Islands	3.02 (1.07, 4.95)	1.42 (1.28, 1.55)	10.70 (9.89, 19.51)
	American Samoa	2.72 (1.79, 2.88)	1.30 (1.29, 1.46)	9.53 (9.41, 13.22)
	Suriname	2.11 (1.68, 2.13)	0.68 (0.45, 1.03)	9.00 (8.56, 9.20)
	Fiji	2.09 (2.07, 2.14)	1.07 (1.04, 2.39)	6.98 (6.00, 7.43)
Central America	Mexico	2.56 (2.10, 3.21)	0.80 (1.11, 2.89)	11.03 (10.84, 11.23)
	Guatemala	2.09 (1.11, 3.47)	0.56 (0.44, 1.23)	9.47 (8.99, 10.62)
	Costa Rica	1.11 (1.03, 1.19)	0.29 (0.21, 0.29)	5.09 (5.00, 5.90)
Caribbean	Dominica	2.37 (2.10, 2.44)	0.76 (0.24, 2.56)	10.16 (9.81,11.44)
	St Vincent & the Grenadines	2.13 (1.85, 2.17)	0.75 (0.69,1.92)	8.76 (6.74,9.00)
	Grenada	1.97 (1.90, 2.41)	0.68 (0.51, 0.79)	8.20 (8.00, 9.20)
	The Bahamas	1.94 (1.94, 3.66)	0.60 (0.44, 1.01)	8.41 (8.31, 10.67)

The rate ratio of non-diabetic CKD mortality to diabetic CKD mortality showed unexpected results (figure 8d). The highest ratios were seen in Central and Eastern European countries, which generally had the lowest SMR for all CKD causes of death; primarily in Albania (7.30:1), Russia (7.28:1), and the United Kingdom (5.22:1). This observation could indicate that the signals detected for “non DM related CKD” are not indicative of CKDu, which could be due to the relatively small number of the population with this mortality outcome.

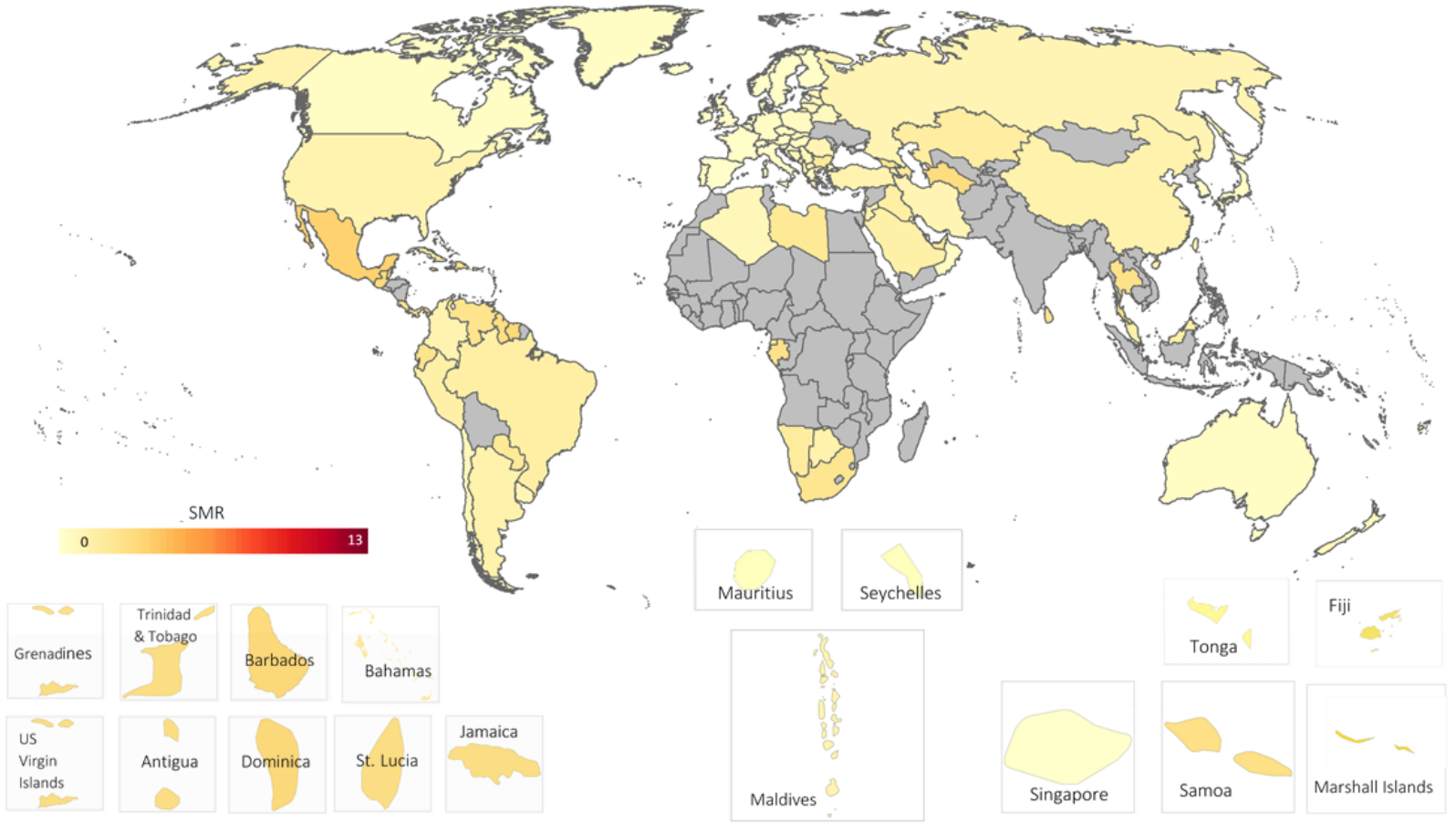


Figure 8a. Spatial distribution of SMR values for all-cause CKD in 15-49-year-old males in higher income countries. Inserts highlight smaller countries

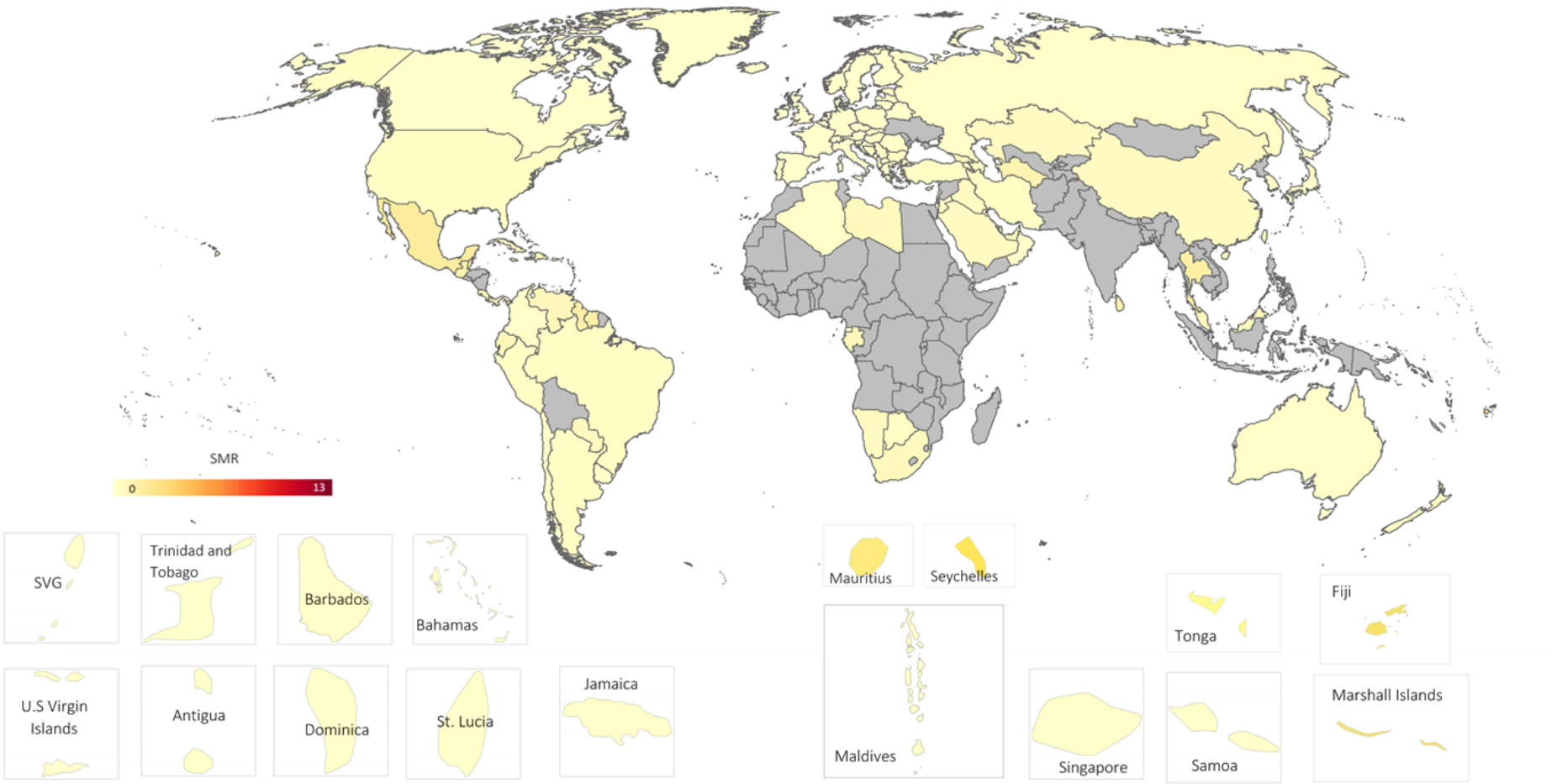


Figure 8b. Spatial distribution of SMR values for CKD attributed to diabetes for 15-49-year-old males in higher income countries. Insert highlights smaller countries

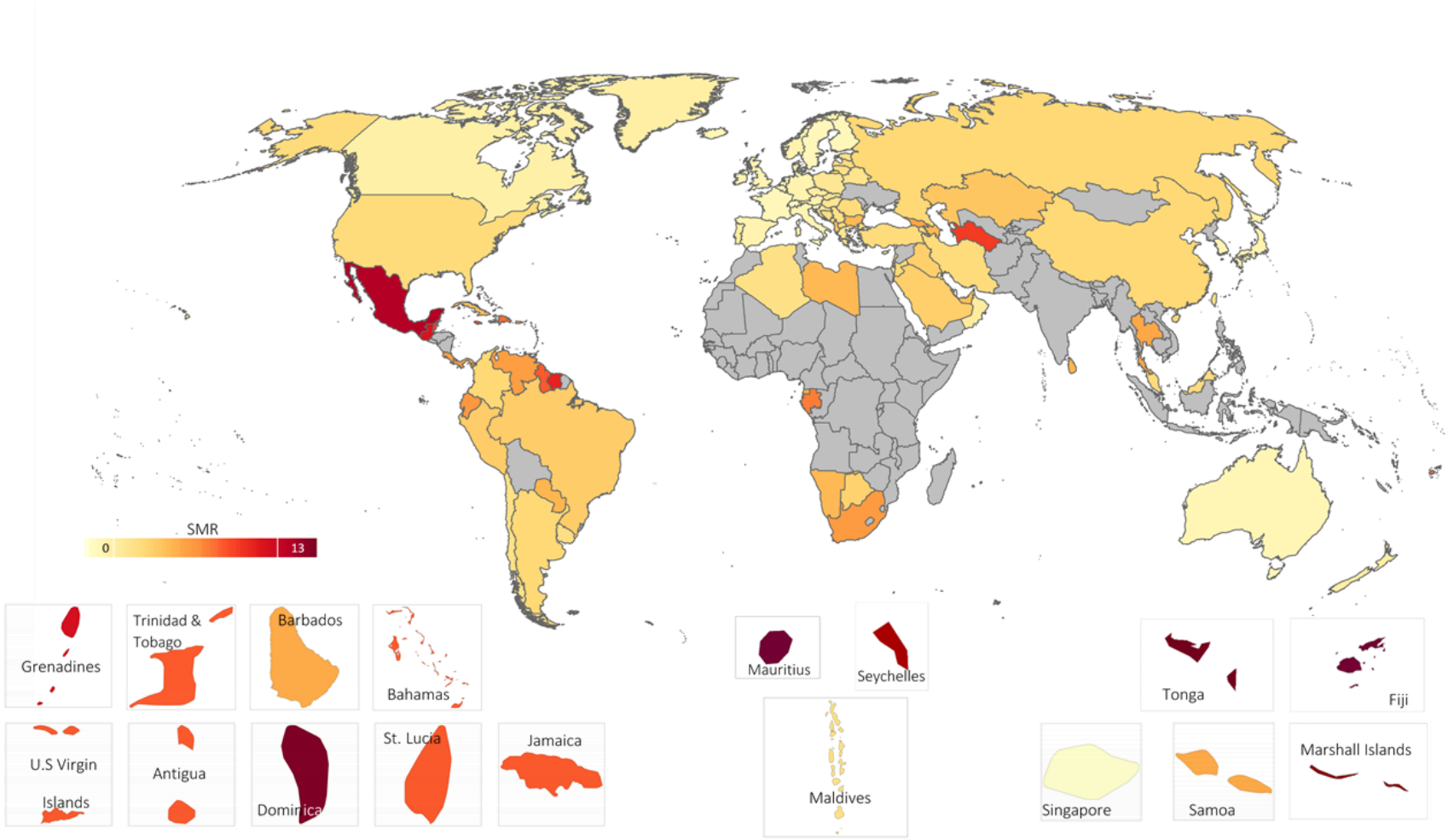


Figure 8c. Spatial distribution of SMR values for non-diabetes related CKD for 15-49-year-old males in higher income countries. Inserts highlight smaller countries

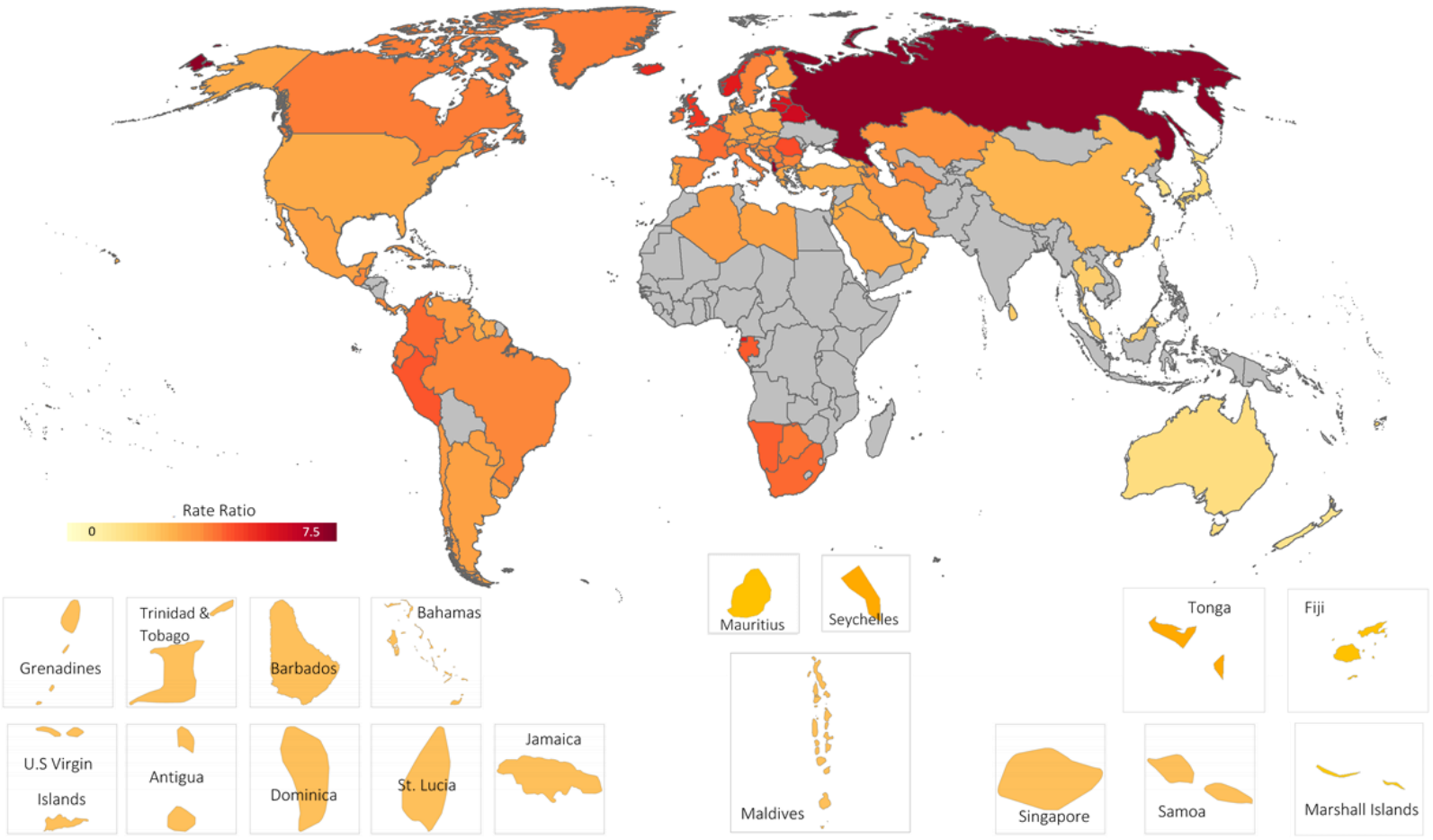


Figure 8d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 15-49-years-old in higher income countries

Males 50-69 years

For this age category, the SMRs for all-cause CKD were generally lower than those observed for the 15-49 males, but the geographical distribution remained the same.

Mauritius had the highest excess mortality (SMR=3.27, 95%CI=3.08, 3.46) followed by Mexico (SMR=2.63, 95%CI=2.61, 2.65) (table 5, figure 9a). Unlike the younger age group, the excess mortality from DM related CKD was generally higher than that observed for-all cause CKD. The highest SMR values were seen in Mauritius (SMR=4.58 95%CI=4.20-4.96) and the Marshall Islands (SMR=2.97 95%CI=1.93-5.01) (table 5, figure 9b). In this age category, the excess deaths for non-DM CKD were the lowest. The highest SMR values were observed in Mauritius (SMR=2.59, CI=2.56-3.04), Mexico (SMR=2.54, 95%CI=1.11-3.48) (table 5, figure 9c).

Table 5. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: "all cause CKD", "CKD attributed to DM" and "non DM related CKD" for males 50-69 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to DM	Non-DM related CKD
			SMR (95% CI)	
Southeast Asia	Mauritius	3.27 (3.08, 3.46)	4.58 (4.20, 4.96)	2.59 (2.56-3.04)
	Seychelles	1.42 (1.34, 1.50)	1.82 (1.70,1.92)	1.21 (1.02,2.50)
	Sri Lanka	0.88 (0.74, 0.99)	1.16 (1.11,1.59)	0.73 (0.41,0.86)
Oceania	Marshall Islands	2.29 (1.24, 3.34)	2.97 (1.93,5.01)	1.93 (1.72,2.01)
	American Samoa	2.18 (1.67, 2.44)	2.83 (1.49,3.00)	1.84 (1.20,2.03)
	Tonga	1.66 (1.60,2.91)	2.21 (2.20,2.43)	1.38 (1.29,2.30)
Central America	Mexico	2.63 (2.61,2.65)	2.81 (2.77,2.84)	2.54 (1.11,3.48)
	Guatemala	1.81 (1.20,1.89)	1.89 (1.70,2.00)	1.76 (1.64,2.58)
	Costa Rica	1.07 (1.01,1.29)	0.95 (0.74,1.02)	1.12 (1.10,1.20)
Latin America	Venezuela	1.43 (1.40,2.00)	1.42 (1.40,2.68)	1.45 (1.20,1.64)
	Ecuador	1.30 (1.29,2.41)	1.22 (1.20,4.08)	1.33 (1.30,2.56)
Caribbean	US Virgin Islands	1.41 (1.10,1.95)	1.52 (1.44,1.69)	1.35 (1.11,2.34)
	Grenada	1.40 (1.20,1.90)	1.57 (1.23,3.68)	1.31 (1.25,2.67)

	Belize	1.28 (1.25,1.30)	1.41 (1.20,4.11)	1.21 (1.10,1.36)

The ratio of non-DM CKD to DM CKD was again concentrated in countries which had the lower SMR values for both “CKD attributed to DM” an “non-DM related CKD”, specifically Russia (5.07:1), Albania (4.44:1) and Norway (3.43:1) (figure 9d)

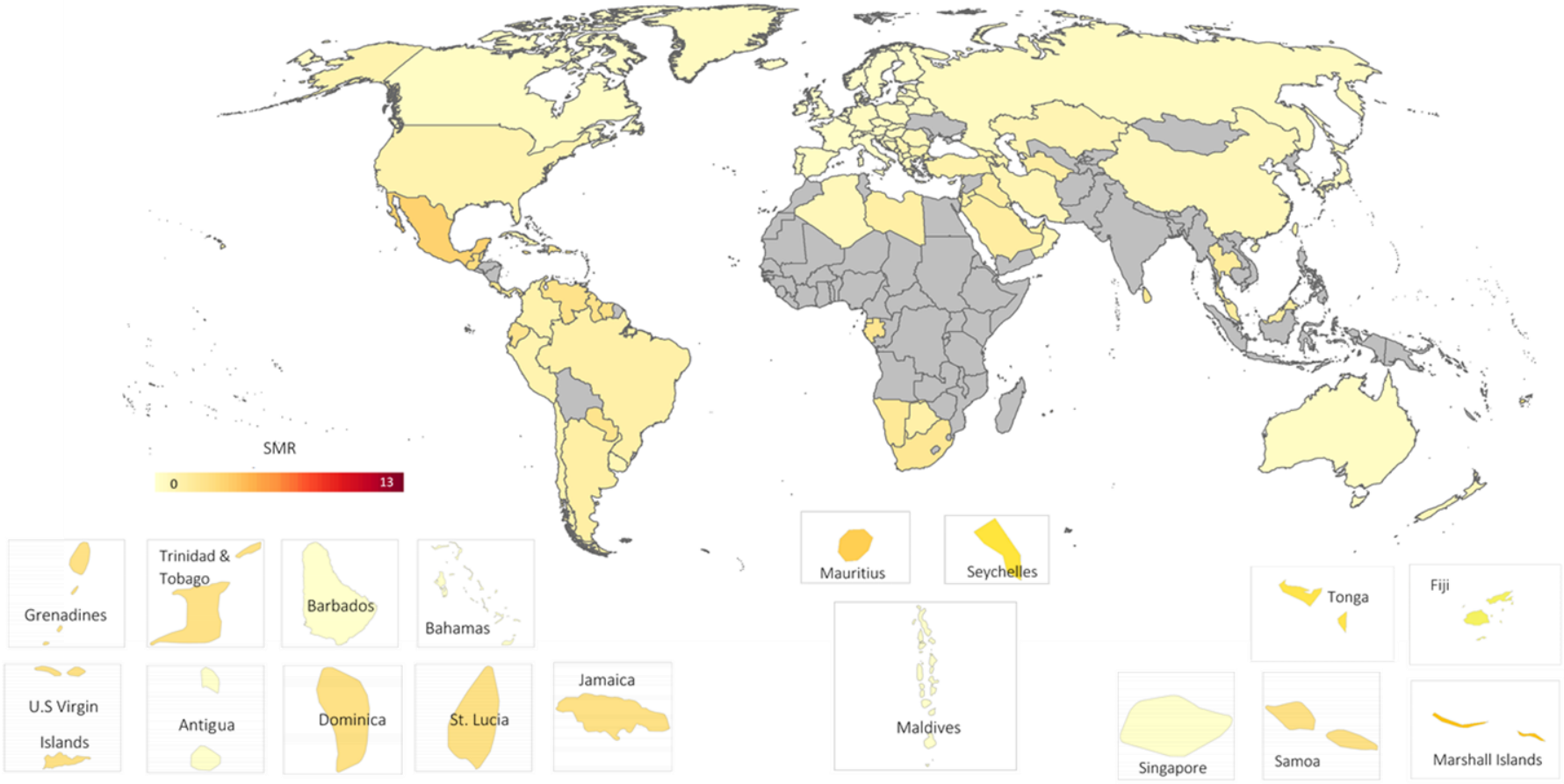


Figure 9a. Spatial distribution of SMR values for all-cause CKD for 50-69-year-old males in higher income countries. Inserts highlight smaller countries

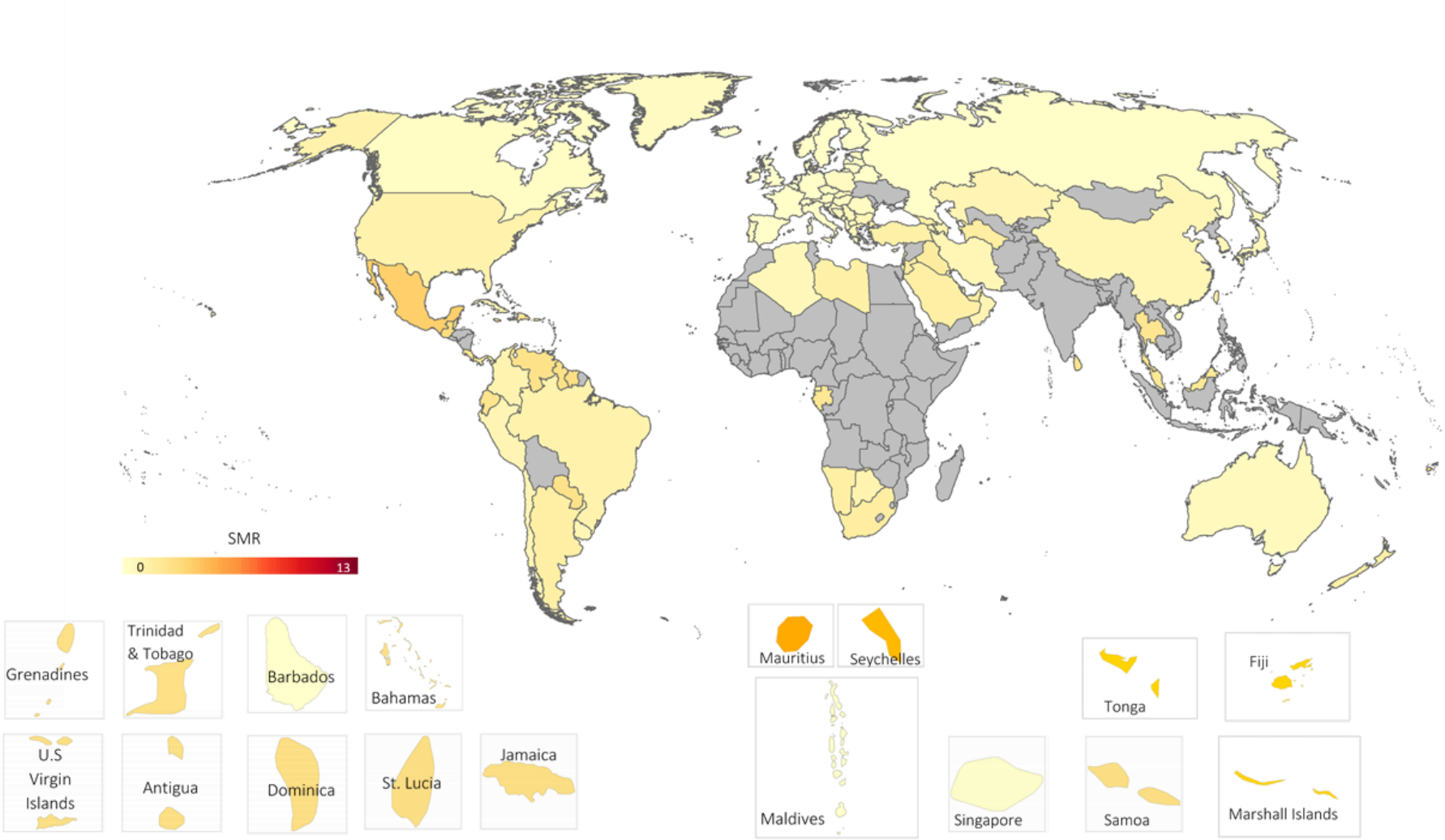


Figure 9b. Spatial distribution of SMR values for CKD attributed to diabetes CKD for 50-69-year-old males in higher income countries. Inserts highlight smaller countries

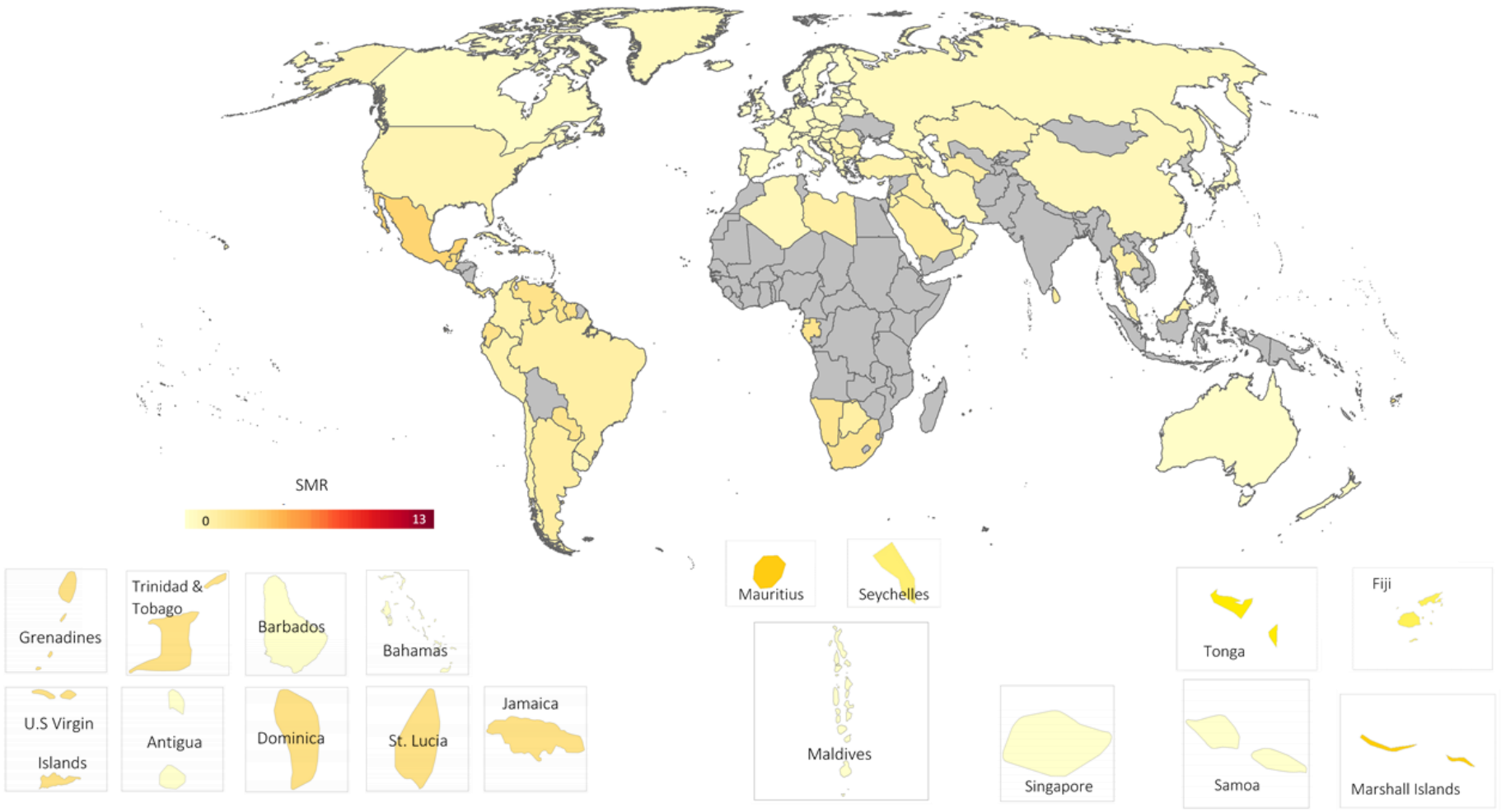


Figure 9c. Spatial distribution of SMR values for Non-diabetes related CKD for 50-69-year-old males in higher income countries. Inserts highlight smaller countries

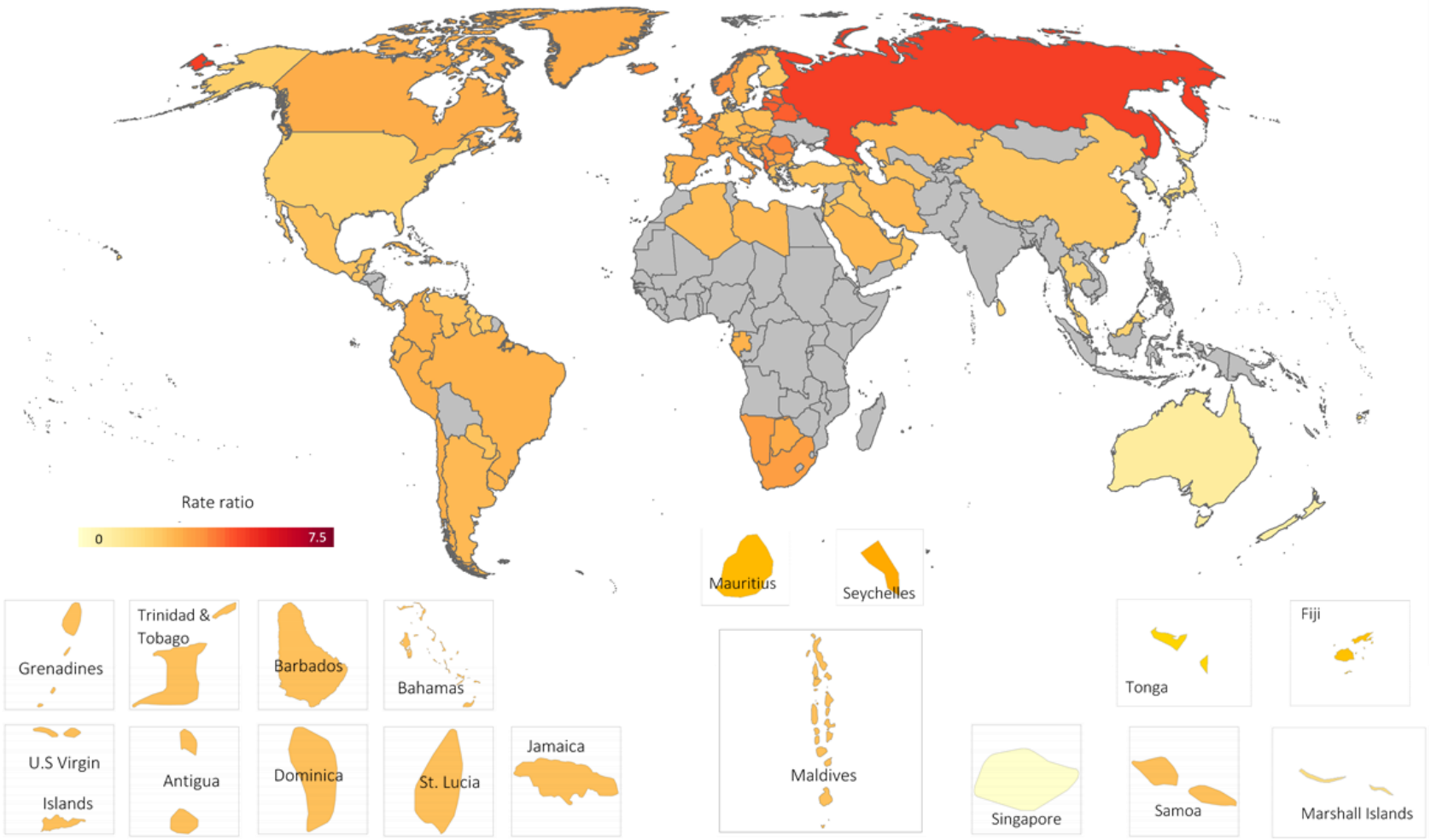


Figure 9d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 50-69-year-old in higher income countries

Males 70+ years

For males, the lowest excess deaths were observed in the 70+ years category. Mauritius had the highest excess mortality for all-cause CKD (SMR=1.42, 95%CI=1.32-1.52), and the lowest was observed in Mexico (SMR=1.14, 95%CI=1.13-1.15) (table 6, figure 10a). As expected for an older age group, the excess deaths for DM CKD were markedly higher than the other mortality categories, with the highest excess deaths seen in Mauritius (SMR=2.18, 95%CI=1.95-2.40) (table 6, figure 10b). For "Non-DM-related CKD", Mauritius was the only country which had excess deaths higher than the global reference (SMR=1.19, 95%CI=1.01-1.22) (table 6, figure 10c).

Table 6. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death :"all cause CKD", "CKD attributed to DM" and "non DM related CKD" for males 70+ years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to DM	Non-DM related CKD
			SMR (95% CI)	
Southeast Asia	Mauritius	1.42 (1.32,1.52)	2.18 (1.95,2.40)	1.19 (1.01,1.22)
Oceania	American Samoa	1.22 (1.04,1.80)	1.76 (1.46,3.06)	1.02 (1.01,2.03)
Central America	Mexico	1.14 (1.13,1.15)	1.29 (1.28,1.31)	1.08 (1.02,1.10)
	Guatemala	1.06 (1.02,1.56)	1.14 (1.02,2.45)	1.02 (1.01,2.36)

The mortality rate ratios again showed the highest non-DM-CKD: DM-CKD ratios are in Russia (6.27:1) and Albania (5.71:1), Iceland (5.47:1), Canada (4.63:1) (figure 10d).

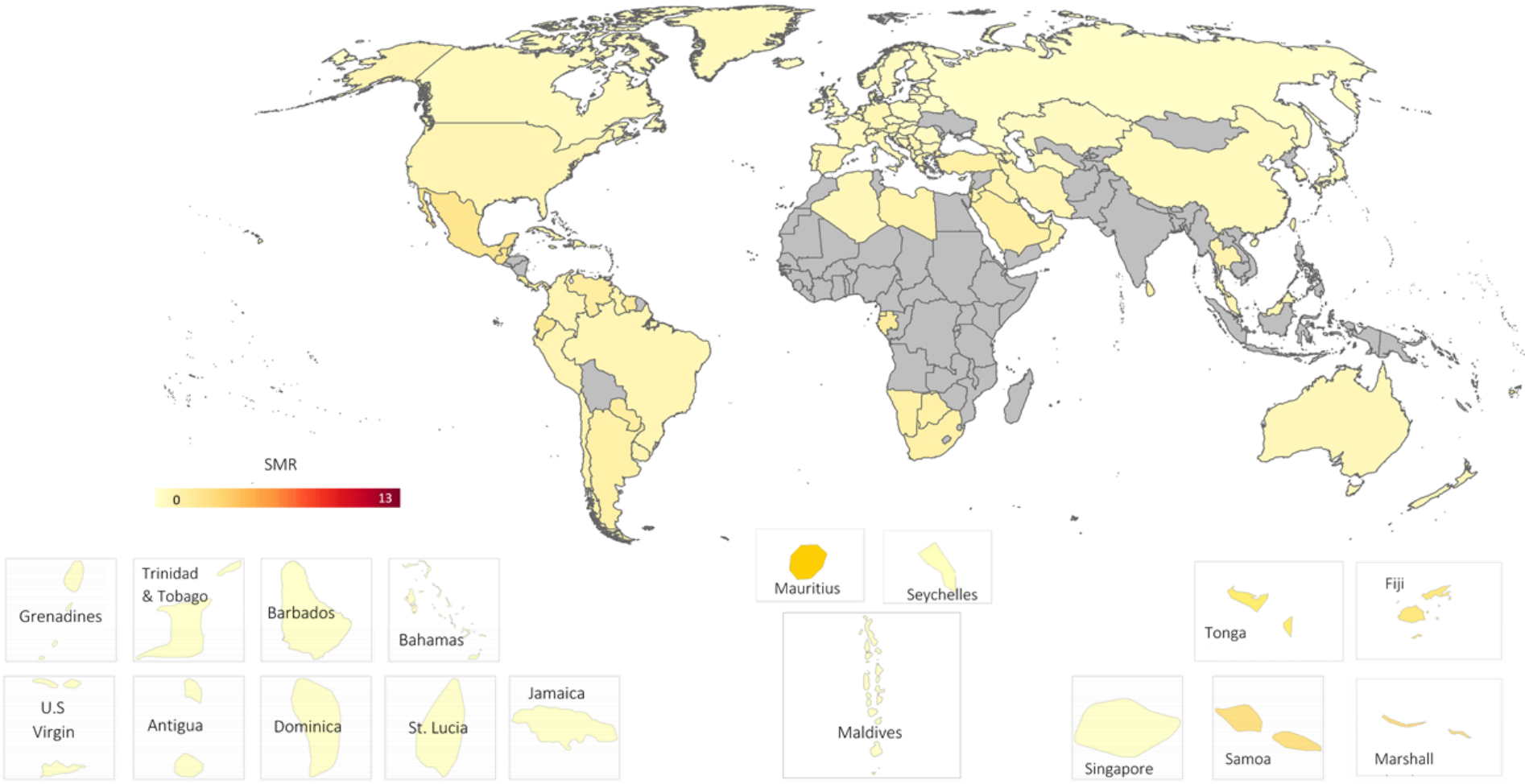


Figure 10a. Spatial distribution of SMR values for all-cause CKD for 70+ year males in higher income countries. Inserts highlight smaller countries

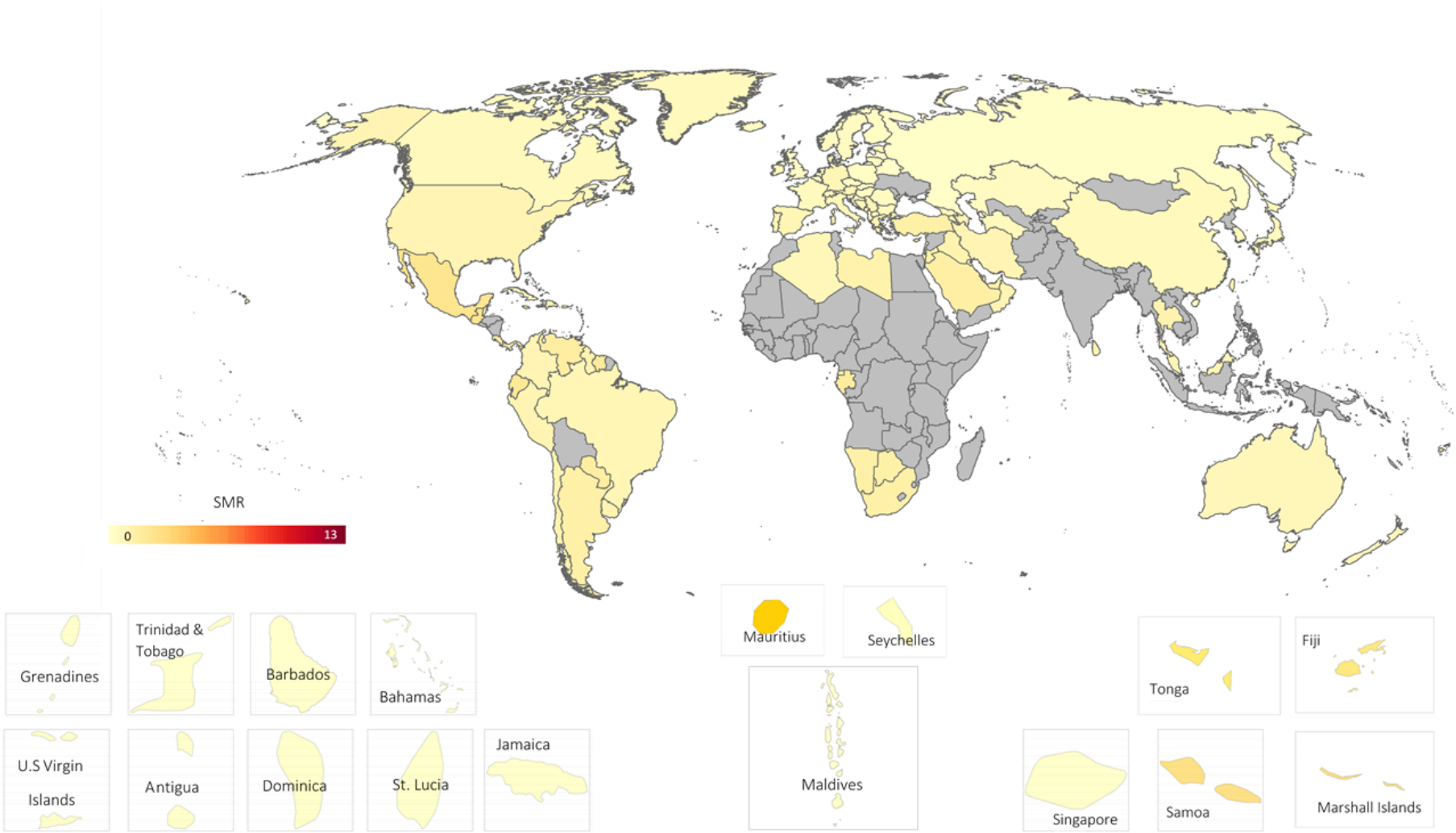


Figure 10b. Spatial distribution of SMR values for CKD attributed to diabetes CKD for 70+ year males in higher income countries. Inserts highlight smaller countries

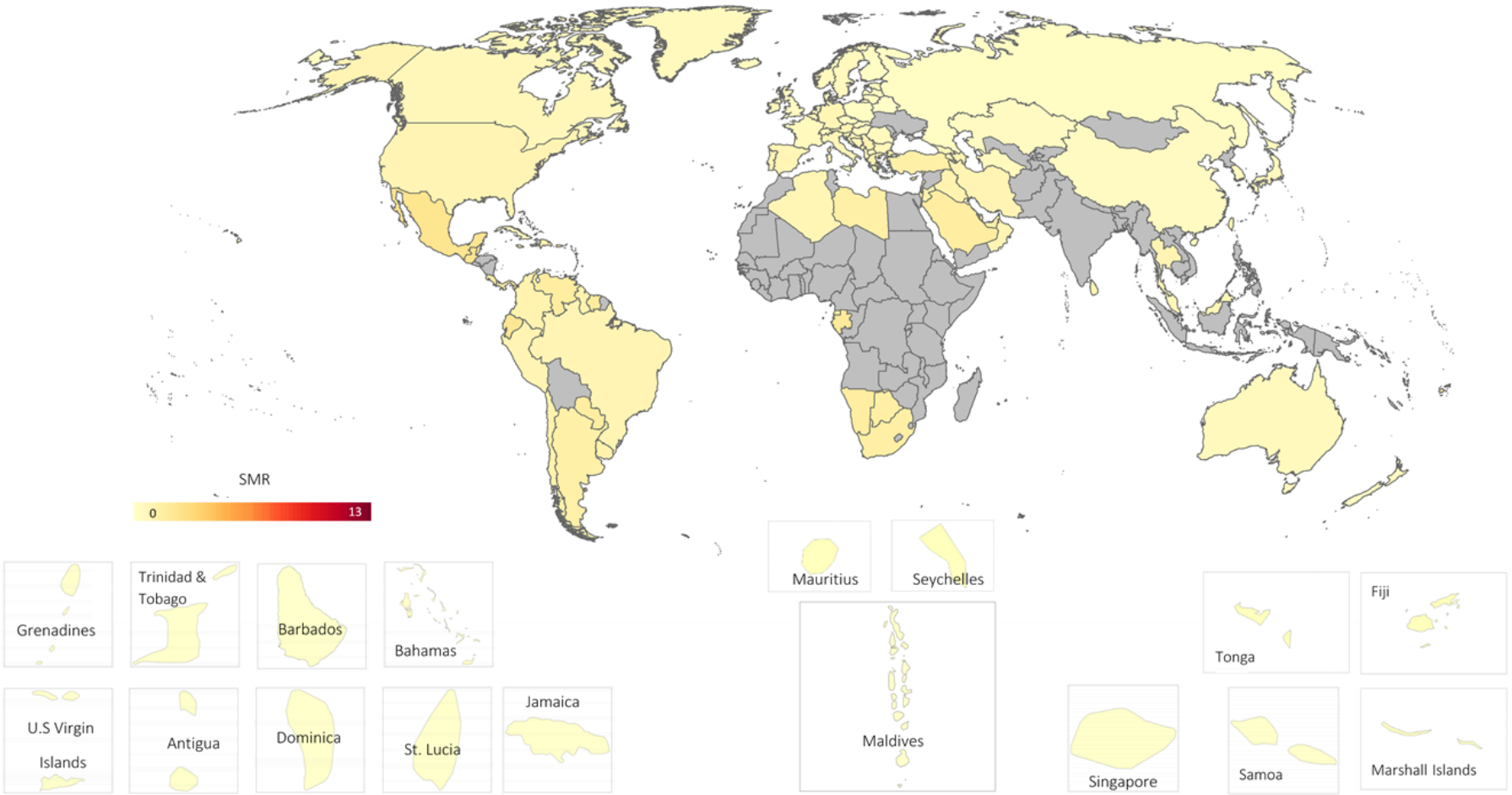


Figure 10c. Spatial distribution of SMR values for non-diabetes related CKD for 70+ year males in higher income countries. Inserts highlight smaller countries.

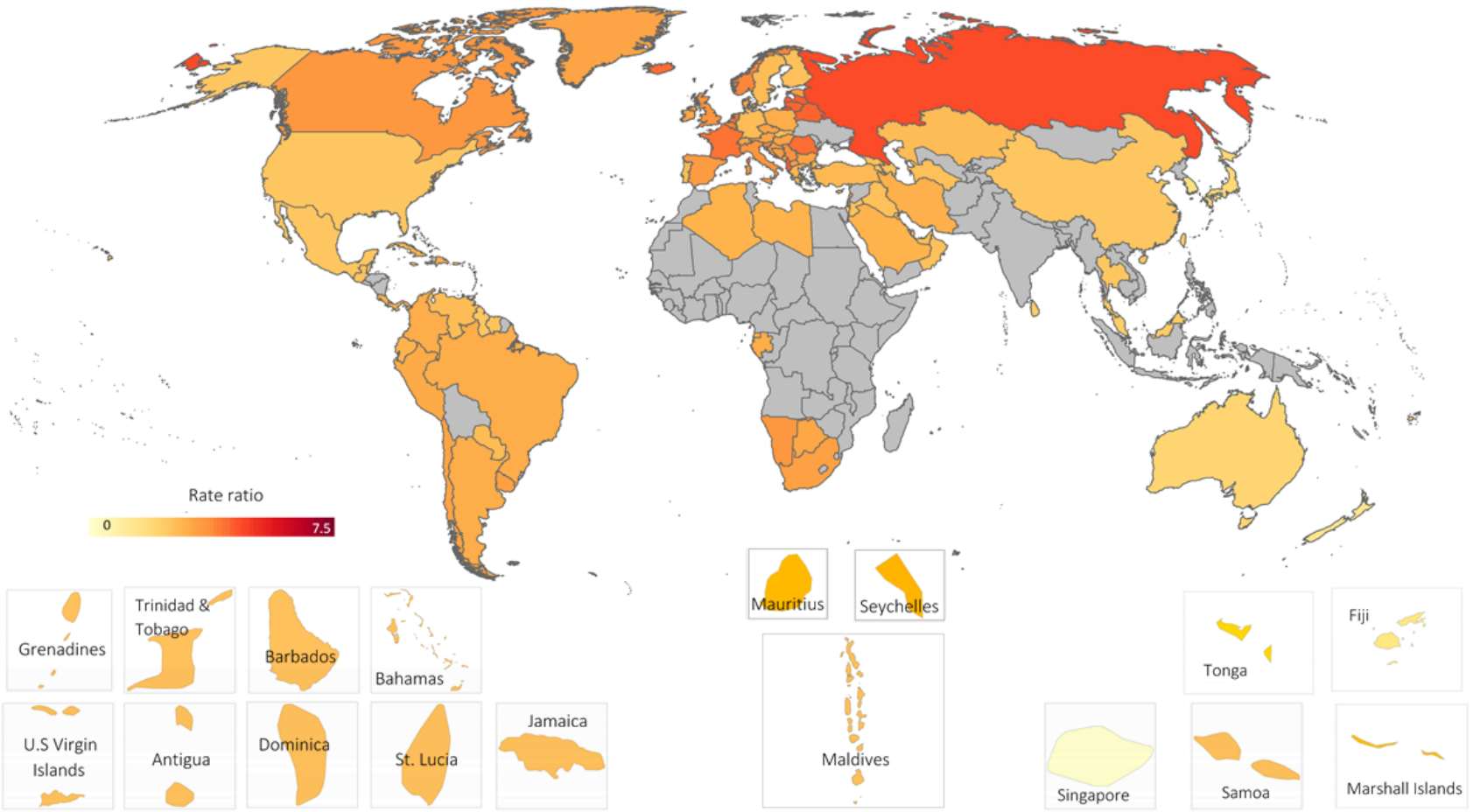


Figure 10d. Spatial distribution of the Mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 70+ years in higher income countries

Females

15-49 years

For females, the geographical distribution of excess deaths did not differ largely from males, with most excess mortality observed across Southeast Asia, Oceania and Central America and the Caribbean. For all cause CKD, the excess deaths in Oceania for females were higher than those of their male counterparts, and the lowest excess deaths were observed in the Caribbean and Central America (table 7, figure 11a). Excess mortality for DM CKD followed a similar pattern (table 7 and figure 11b). In comparison to their male counterparts, excess deaths for non-DM CKD were lower for females. Within this group however, the SMR values were markedly higher than the other CKD mortality causes, specifically in the Marshall Islands (SMR=8.63 95%CI=2.60-9.65) and Dominica (SMR=5.12 95%CI=4.91-5.65) (Table 7 and figure 11c), which could be indicative of a signal of CKDu in this group. The higher mortality rate ratios of non-DM to DM CKD were seen in Lebanon (5.31:1) and Saudi Arabia (5.08:1), the USA (4.66:1) and Russia (4.26) (figure 11d).

Table 7. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death for "all cause CKD", "CKD attributed to DM" and "non DM related CKD" for females 15-49 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to DM	Non-DM related CKD
			SMR (95% CI)	
Oceania	Marshall Islands	6.35 (5.97,6.44)	4.60 (3.99,4.91)	8.63 (2.60,9.65)
	American Samoa	4.22 (4.13,5.69)	3.03 (2.98,3.21)	5.78 (5.69,7.66)
	Samoa	2.78 (2.44,4.97)	2.02 (2.00,2.59)	3.78 (2.11,4.62)
Southeast Asia	Mauritius	3.06 (2.03,4.11)	2.61 (1.11,2.89)	3.64 (2.17,4.22)
Caribbean	Dominica	3.01 (2.55,3.99)	1.38 (1.24,2.56)	5.12 (4.91,5.65)
	St Vincent & the Grenadines	2.43 (2.39,4.33)	1.29 (1.11,2.56)	3.93 (3.22,4.17)
	The Bahamas	2.08 (2.01,3.22)	1.03 (1.02,1.33)	3.46 (3.41,4.00)
	Grenada	2.33 (1.97,2.41)	1.19 (1.11,2.31)	3.81 (2.79,3.88)

Central America	Mexico	2.44 (2.30,4.13)	1.32 (1.31,3.22)	3.89 (3.44,4.61)
	Guatemala	2.01 (1.97,2.44)	0.90 (0.58,0.98)	3.44 (2.59,4.32)

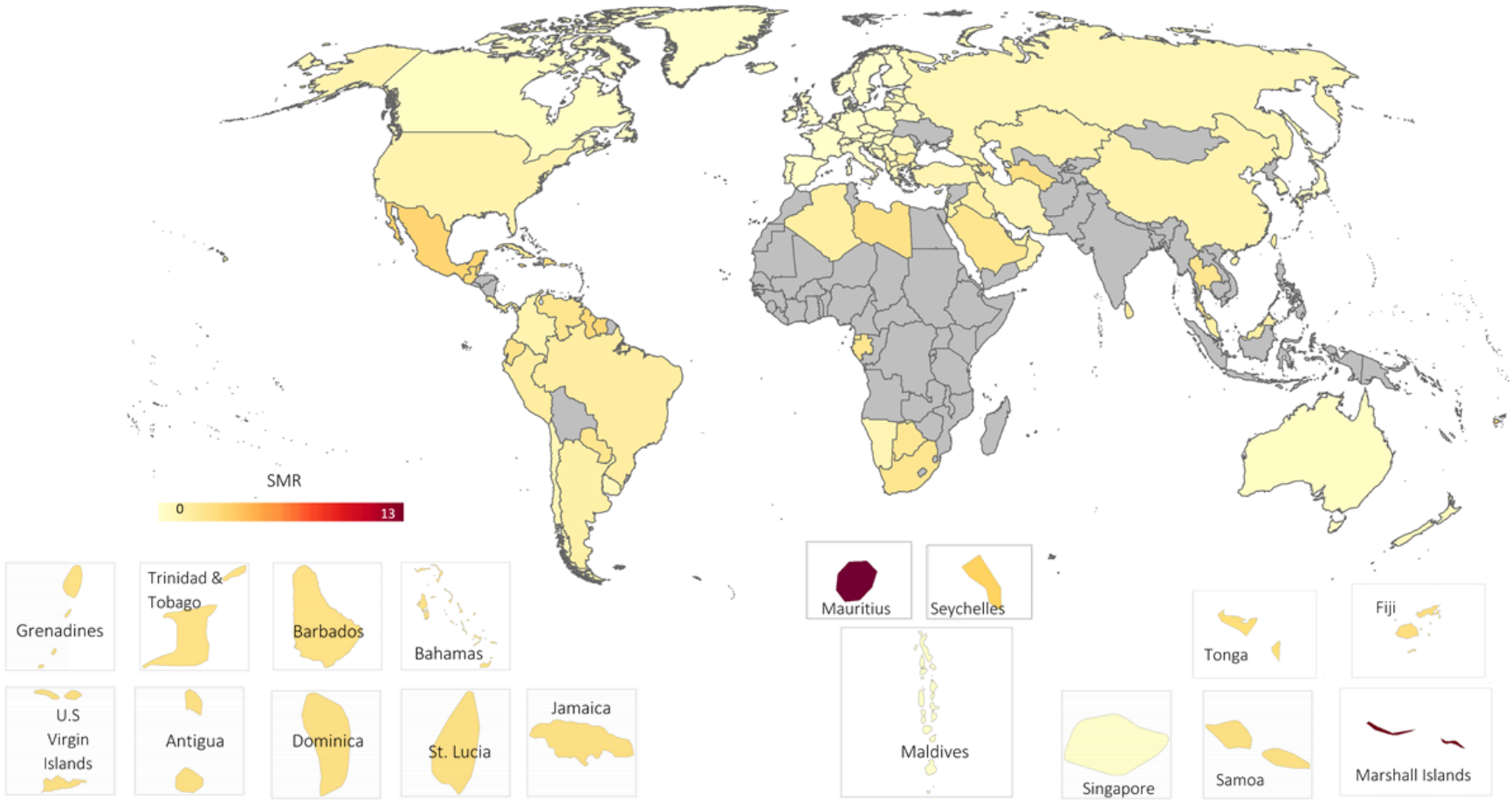


Figure 11a. Spatial distribution of SMR values for all-cause CKD for 15-49-year females in higher income countries. Inserts highlight smaller countries

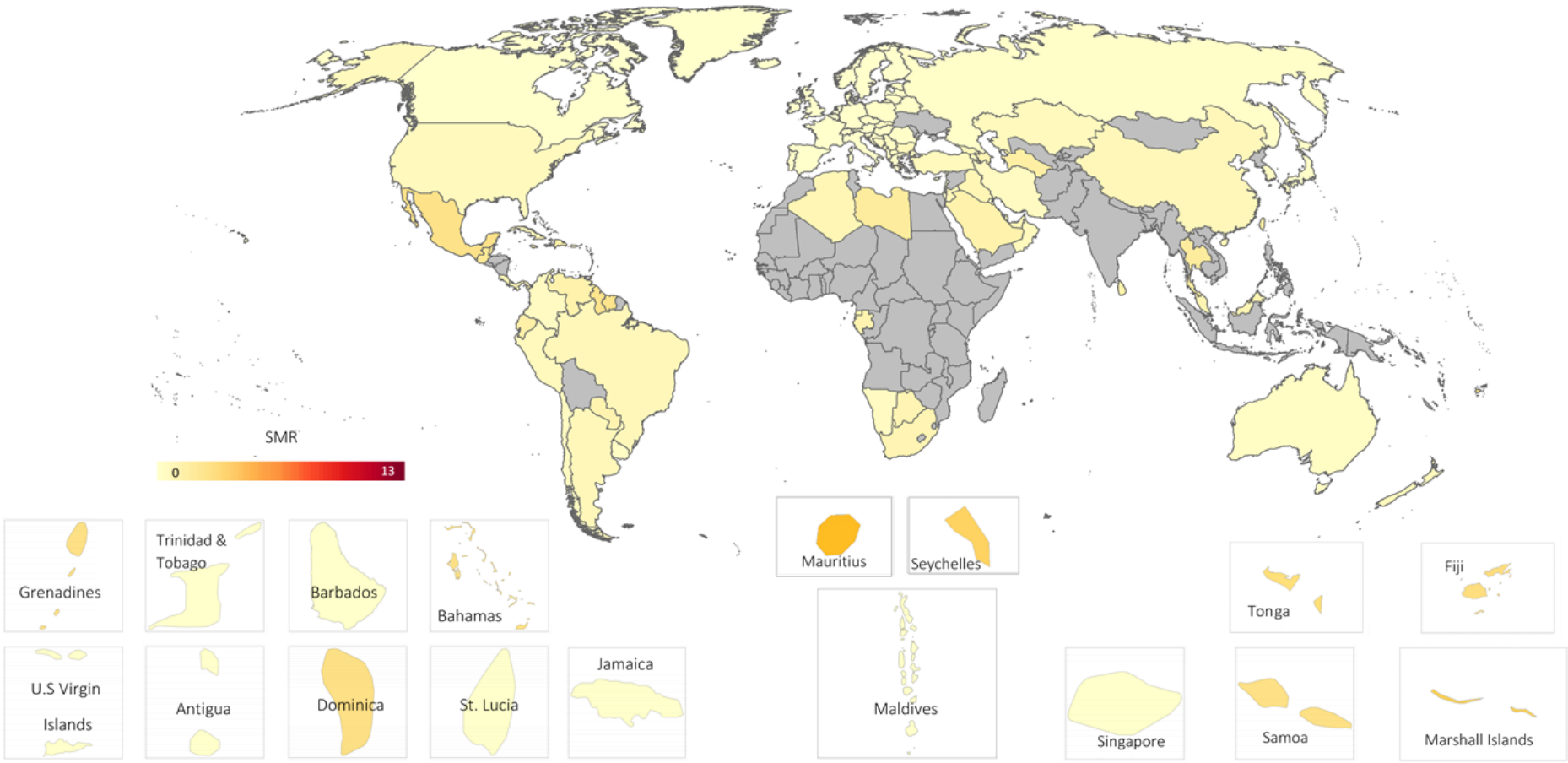


Figure 11b. Spatial distribution of SMR values for CKD attributed to diabetes for 15-49-year females in higher income countries. Inserts highlight smaller countries

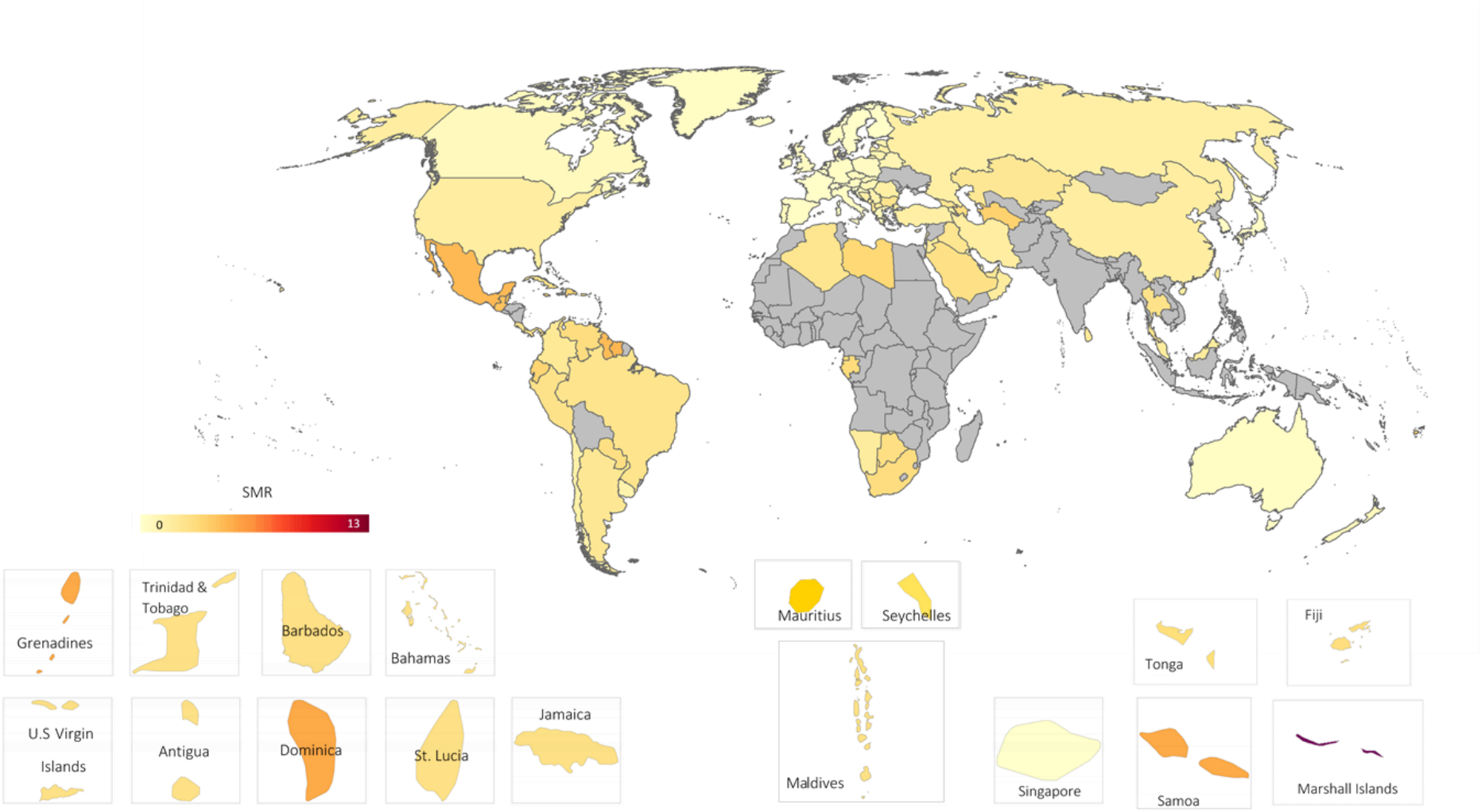


Figure 11c. Spatial distribution of SMR values for non-diabetes related CKD for 15-49-year females in higher income countries. Inserts highlight smaller countries

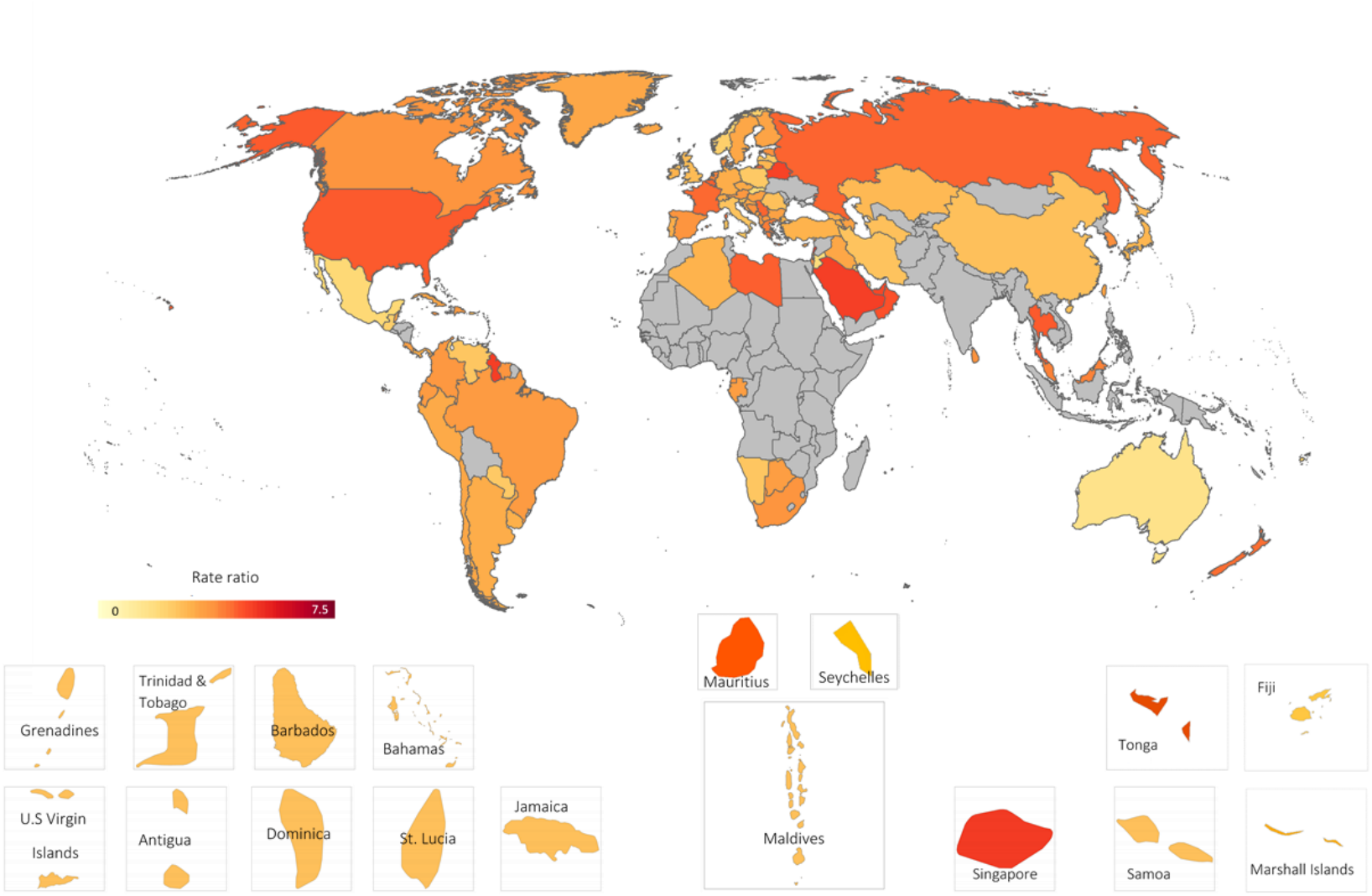


Figure 11d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for females 15-49 years in high income countries

Females 50-69 years

A similar geographical distribution of excess deaths was seen for the 50-69-year category, however, there were also signals of "all cause CKD" detected in South American countries, the highest of which was in Ecuador. In a similar pattern observed for males, the excess deaths for all-cause CKD decreased with age. The highest SMR was observed in the Marshall Islands (SMR=3.84 95%CI=2.06,5.61) however most excess deaths were largely concentrated across central and southern American countries, the highest of which were in Mexico (SMR=3.53 95%CI=3.50,3.56) and Ecuador (SMR=2.13 95%CI=2.06,2.20) (table 8, figure 12a). Excess deaths for DM CKD were higher than those for all cause CKD, with many of the higher values seen across Oceania and Southeast Asia, specifically the Marshall Islands and Mauritius (table 8, figure 12b). Excess mortality for non-DM CKD was the lowest of all causes of mortality. The strongest signal was detected in The Marshall Islands however, most of the high SMR values were clustered in Central American countries, the highest of which was Mexico (table 8, figure 12c).

Table 8. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death :"all cause CKD", "CKD attributed to DM" and "non DM related CKD" for females 50-59 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to DM	Non-DM related CKD
		SMR (95% CI)		
Oceania	Marshall Islands	3.84 (2.06, 5.61)	5.10 (1.64,8.57)	3.15 (1.15,5.15)
	American Samoa	2.76 (2.40,3.41)	3.65 (3.00,4.79)	2.28 (1.82,3.73)
	Samoa	2.35 (2.01,3.45)	3.14 (2.36,4.51)	1.93 (1.11, 2.75)
	Tonga	2.32 (2.24,4.56)	3.16 (3.00,4.26)	1.88 (1.79,2.56)
Central America	Mexico	3.53 (3.50,3.56)	4.09 (4.03,4.14)	3.23 (3.19,3.26)
	Guatemala	2.60 (2.51,5.69)	3.02 (2.50,3.67)	2.37 (2.22,4.53)
Southeast Asia	Mauritius	3.41 (3.23,4.57)	4.84 (4.77,5.23)	2.64 (2.59,3.52)
	Seychelles	1.73 (1.61,1.78)	2.26 (2.21,3.45)	1.45 (1.34,2.36)

Latin America	Ecuador	2.13 (2.06,2.20)	2.27 (2.16,2.59)	2.05 (1.91,3.98)
	Venezuela	1.53 (1.49,2.00)	1.71 (1.60,2.34)	1.43 (1.40,3.22)
Caribbean	Belize	1.92 (1.45,3.01)	2.24 (2.14,3.46)	1.74 (1.60,2.21)
	Dominica	1.66 (1.55,1.70)	1.94 (1.10,2.35)	1.50 (1.43,3.45)

The highest mortality rate ratios of non-diabetes-related CKD to diabetes-related-CKD were in Russia (3.91:1) and Latvia (3.46:1), and Lithuania (3.45) (figure 12d).

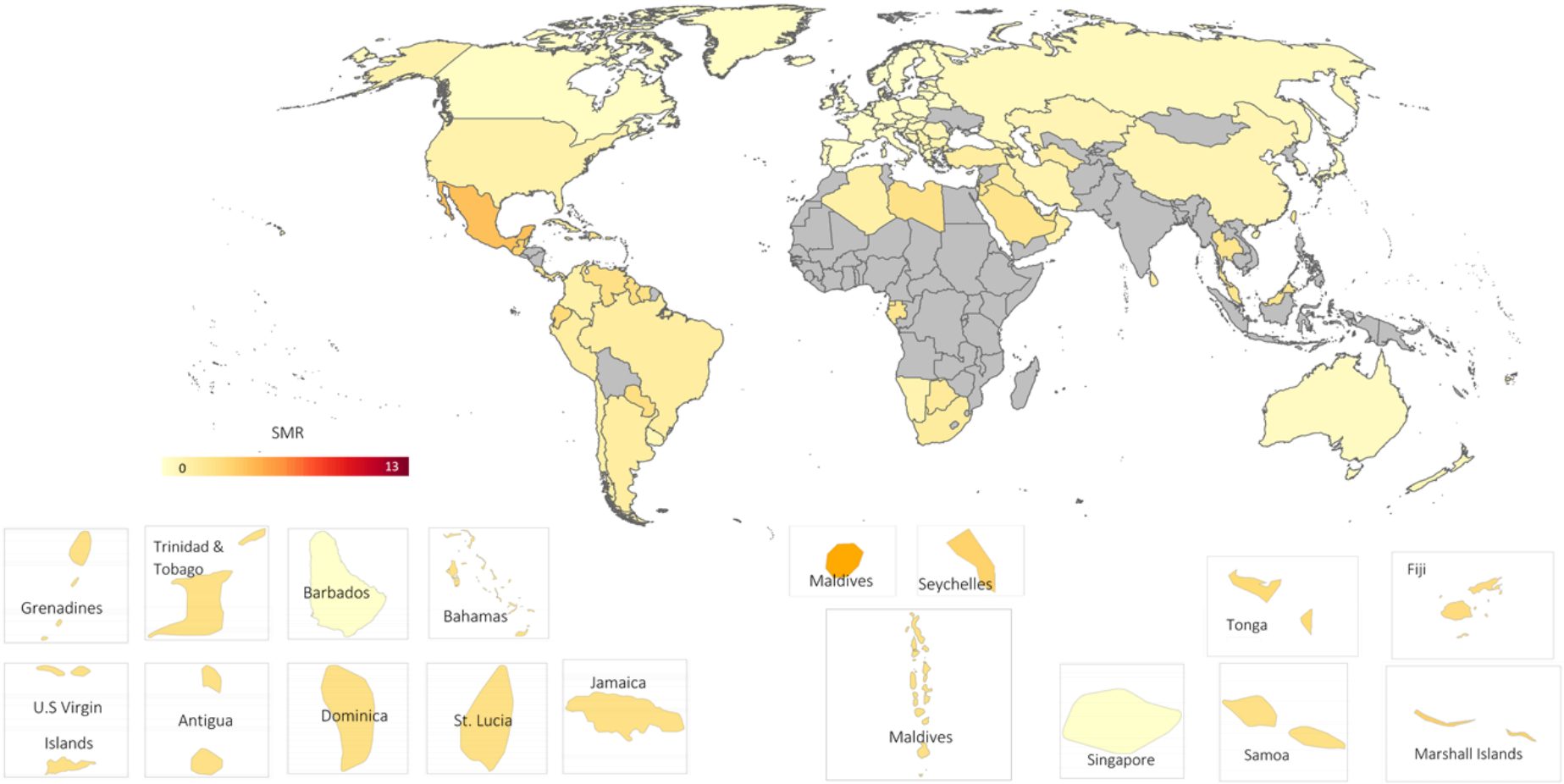


Figure 12a. Spatial distribution of SMR values for all-cause CKD for 50-69-year females in higher income countries. Inserts highlight smaller countries

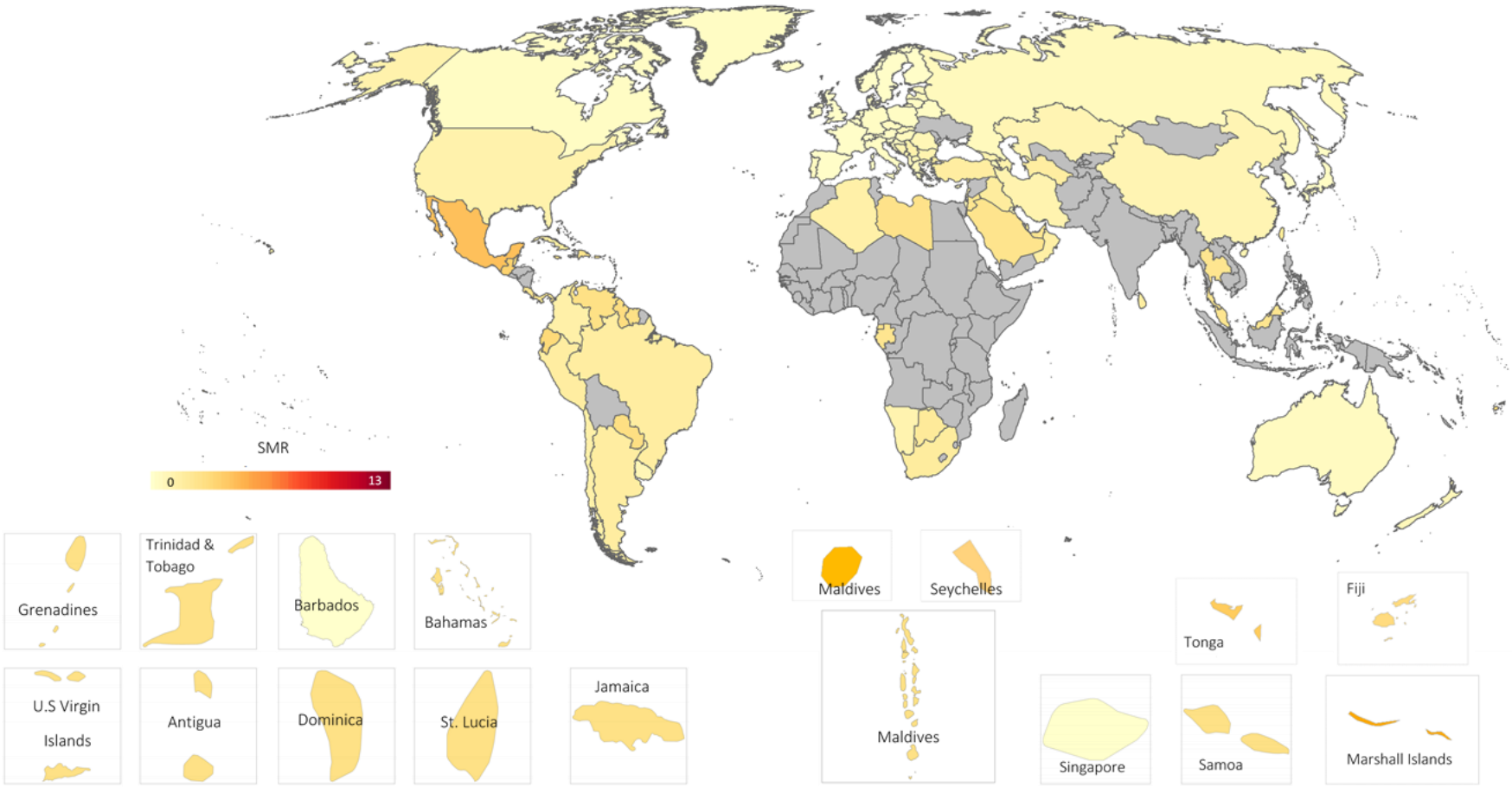


Figure 12b. Spatial distribution of SMR values for CKD attributed to diabetes for 50-69-year females in higher income countries. Inserts highlight smaller countries

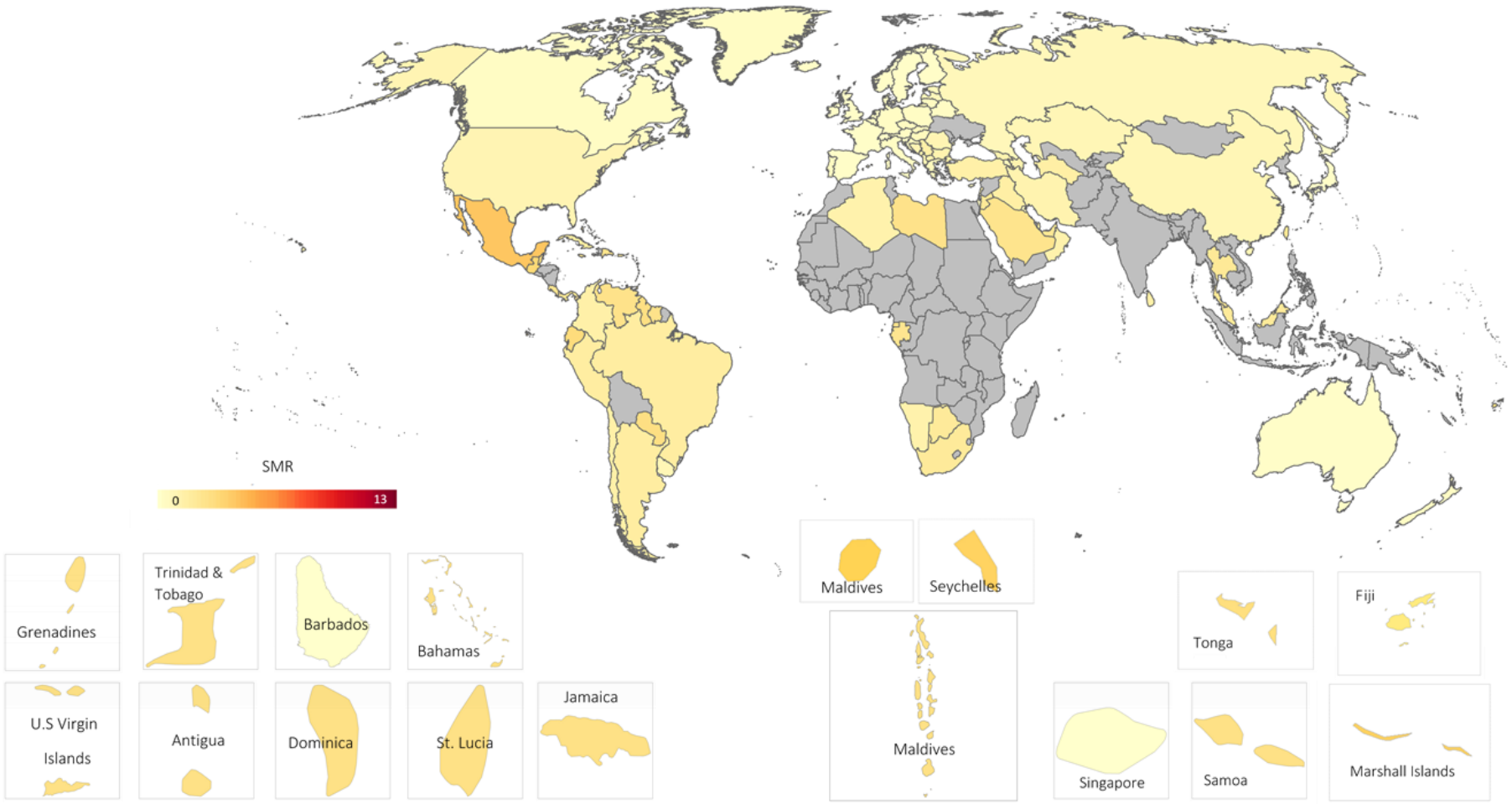


Figure 12c. Spatial distribution of SMR values for non-diabetes related CKD for 50-69-year females in higher income countries. Inserts highlight smaller countries

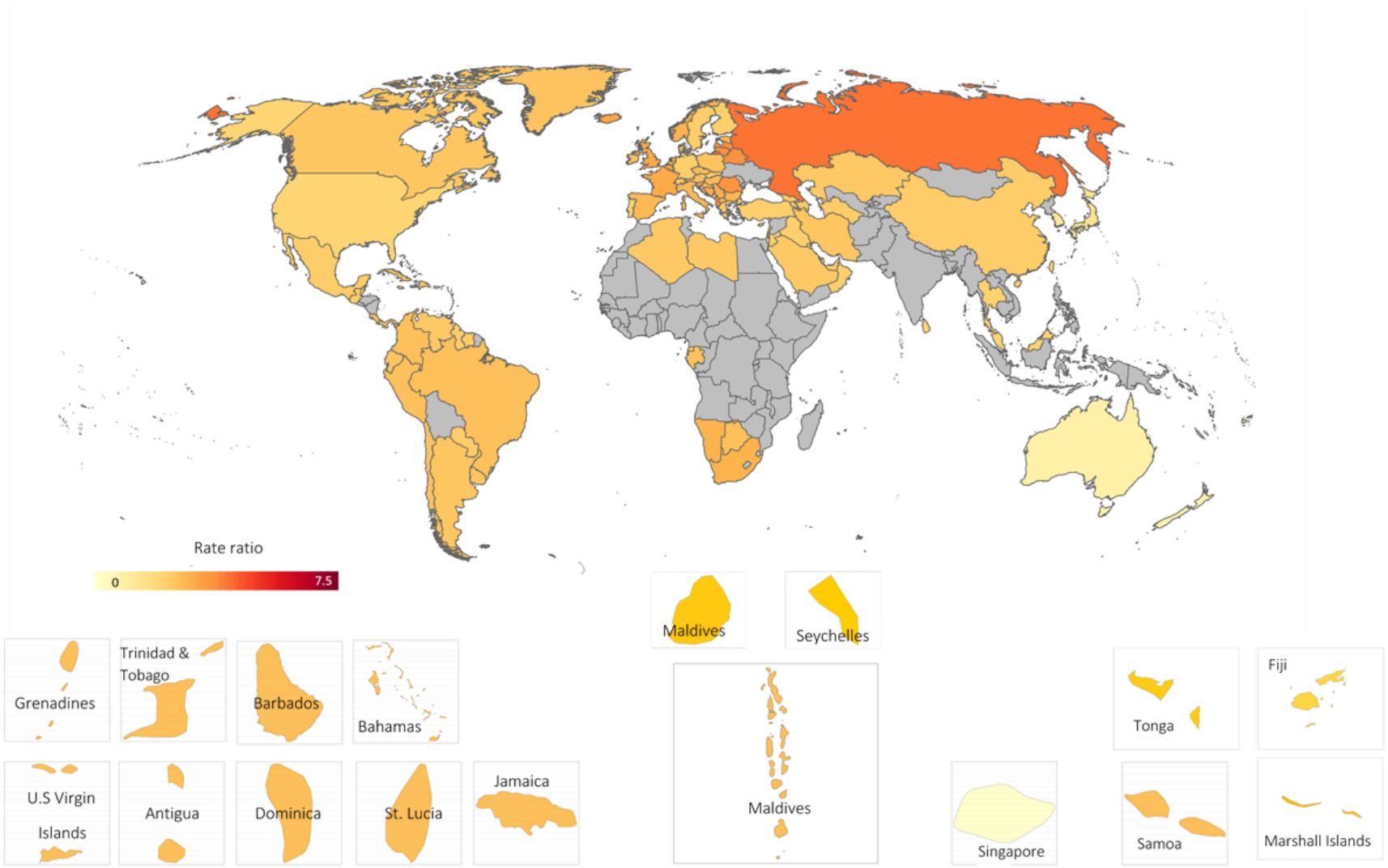


Figure 12d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for females 50-69 years in higher income countries. Inserts highlight smaller countries

Females 70+ years

Like their male counterparts, the lowest excess mortality for all causes of death were observed in the 70+ age group. The geographical distribution remained the same, however in Oceania, American Samoa had the highest excess deaths for the region.

For all cause CKD, excess mortality was highest in American Samoa (SMR=2.10 95%CI=1.13-3.08) (table 9, figure 13a). With the exception of Ecuador, excess deaths for DM CKD were higher than those for all cause CKD (table 1) (table 9, figure 13b). For non-DM CKD the highest excess deaths remained in Mauritius (SMR=3.23, CI=3.21-4.66), and were higher than those for the other regions (table 9, figure 13c).

Table 9. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death:"all cause CKD", "CKD attributed to DM" and "non DM related CKD" for females 70+ years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to DM	Non-DM related CKD
			SMR (95% CI)	
Oceania	American Samoa	2.10 (1.13,3.08)	2.81 (2.44,3.09)	1.78 (1.45,2.11)
	Tonga	1.96 (1.34,2.97)	3.16 (3.00,4.34)	1.88 (1.80,2.34)
	Marshall Islands	1.81 (1.78,1.90)	5.10 (3.98,6.11)	3.15 (3.02,4.55)
Southeast Asia	Mauritius	2.08 (1.90,2.36)	2.95 (2.06,3.15)	3.23 (3.21,4.66)
	Seychelles	1.71 (1.67,3.44)	2.26 (2.21,3.45)	1.45 (1.10,3.22)
Central America	Mexico	1.93 (1.90,2.17)	2.32 (1.91,2.94)	1.68 (1.01,1.79)
	Guatemala	1.41 (1.32,1.76)	3.02 (3.00,3.56)	2.37 (1.68,3.45)
South America	Ecuador	1.64 (1.59,2.03)	1.59 (1.54,2.26)	1.66 (1.41,3.92)

For this age category the highest mortality rate ratios were in Russia (4.73:1), Albania (4.35:1), and Canada (3.12:1) (figure 13d).

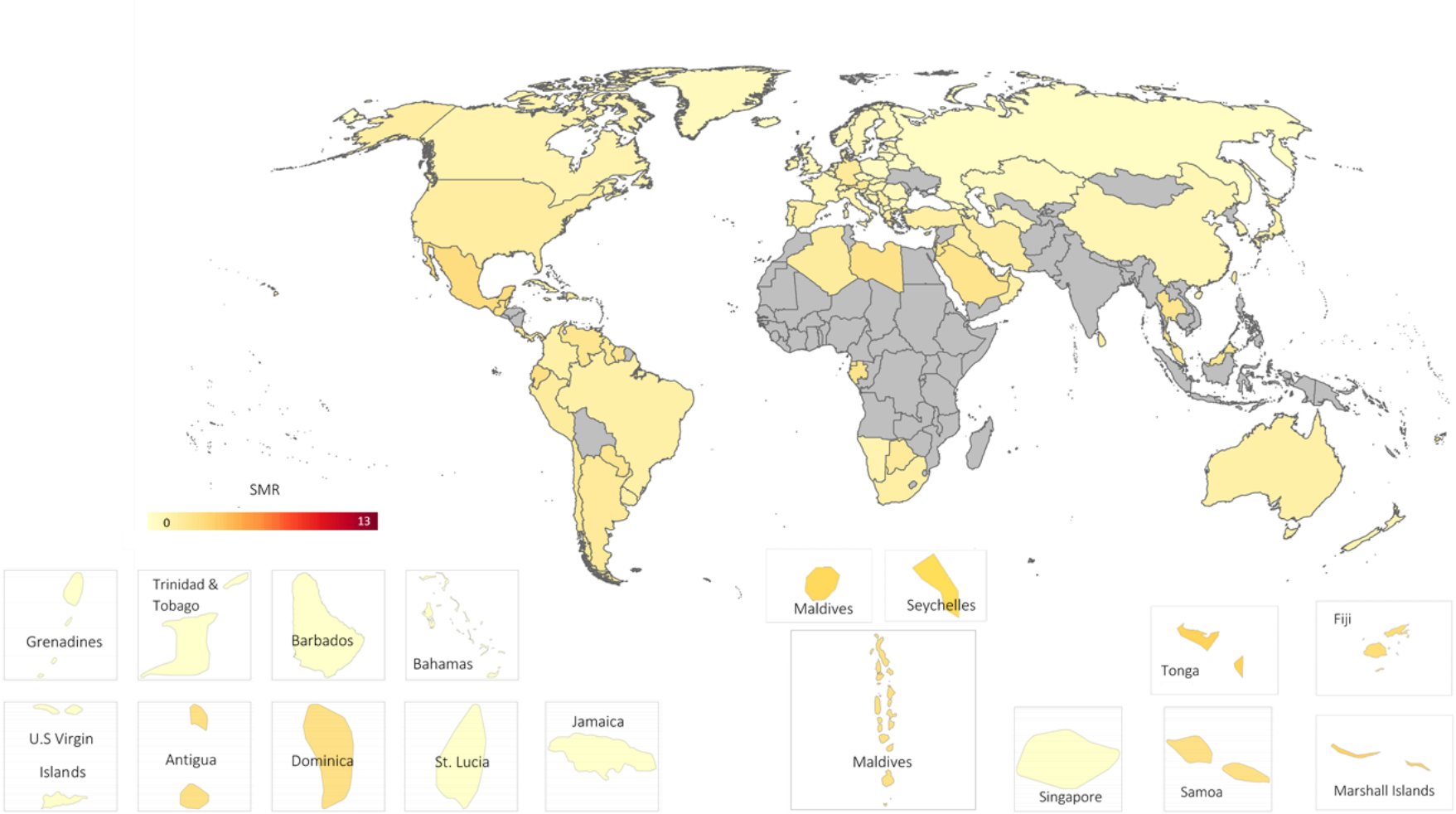


Figure 13a. Spatial distribution of SMR values for all-cause CKD for 70+ year females in higher income countries. Inserts highlight smaller countries

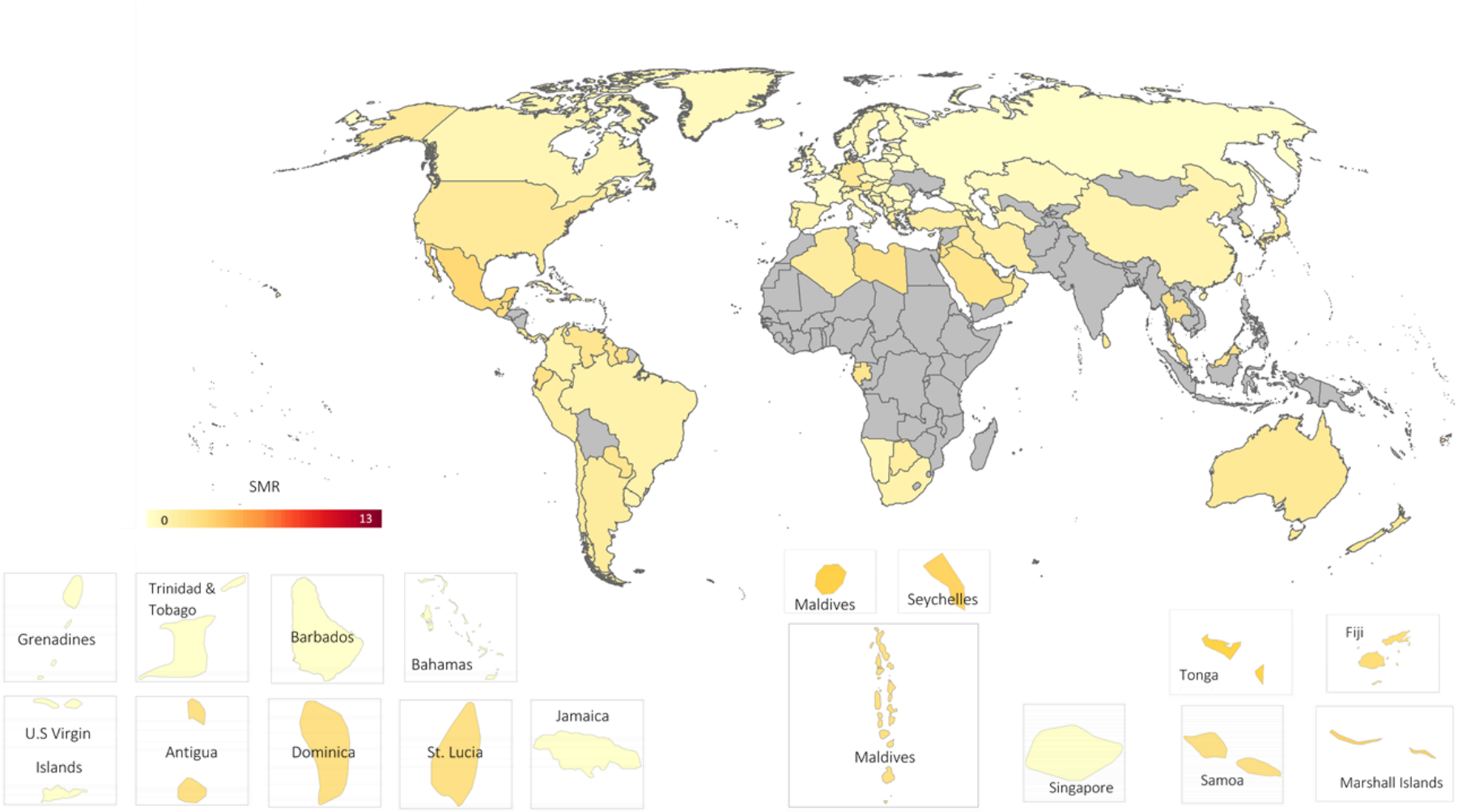


Figure 13b. Spatial distribution of SMR values for CKD attributed to diabetes for 70+ year females in higher income countries. Inserts highlight smaller countries

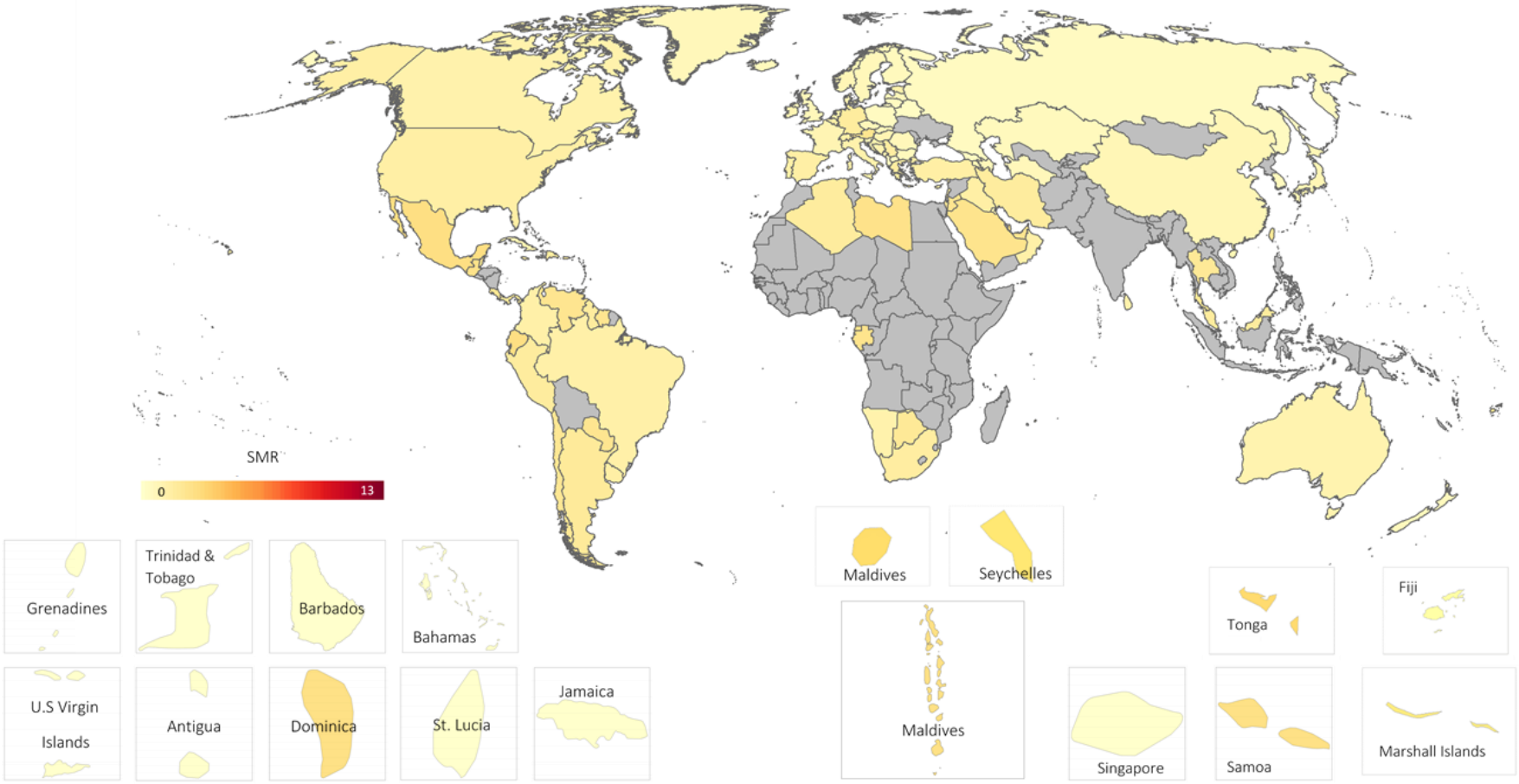


Figure 13c. Spatial distribution of SMR values for non-diabetes related CKD for 70+ year females in higher income countries. Inserts highlight smaller countries

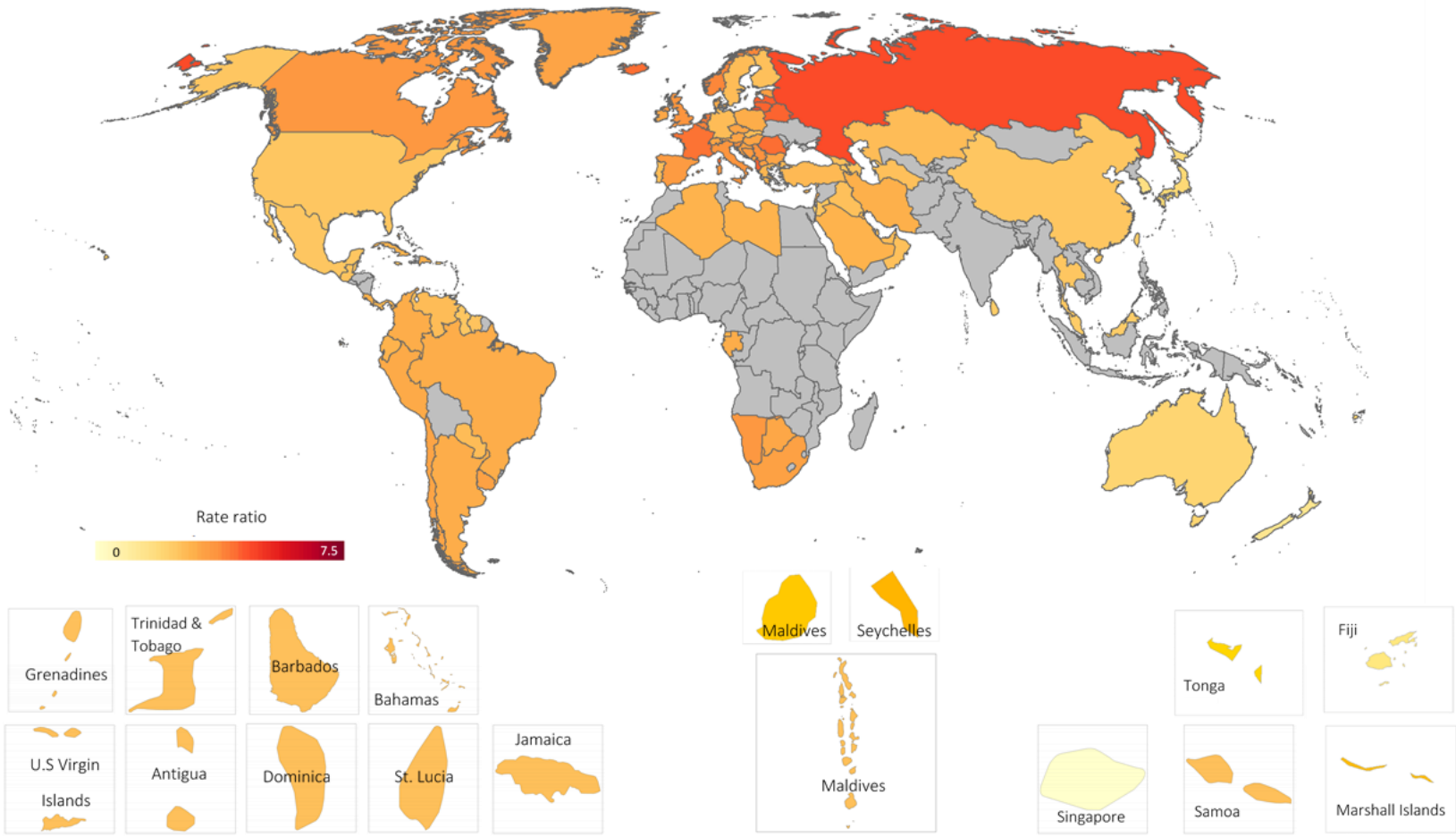


Figure 13d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for females 70+ years in higher income countries. Inserts highlight smaller countries

3.3.3 Low-income countries

Males

15-49 years

For lower income males, mortality from all-cause CKD was higher than those of their higher-income counterparts. The geographical distribution of excess deaths was generallyseen across Oceania and Central-Southern America, in addition across Sub-Saharan Africa (SSA).

Central America had the highest overall excess mortality for all cause CKD. El Salvador (SMR=5.62, CI=5.35, 5.89) had a notably higher SMR than other countries. In SSA there were many regions which had high excess mortality, most notably in eSwatini (SMR=3.10, 95%CI=3.00,3.59) (table 10, figure 14a). For DM CKD excess mortality was the lowest across all regions. The highest SMR was in El Salvador (SMR=1.60, CI=1.44, 1.75), and the remainder of excess deaths were centred in Oceania in Papua New Guinea (SMR=1.52, CI=1.40, 1.64) and the Solomon Islands (SMR=1.51, CI=1.50, 2.34). There were no excess deaths recorded in Sub-Saharan Africa for this category. (table 8, figure 14b). For non-DM related CKD El Salvador and Nicaragua had the highest excess mortality across age and sex in these analyses (SMR=25.03, CI=23.67,26.39, and SMR= 16.66 (16.07, 17.69 respectively). There were also signals detected across SSA, the highest of which was in eSwatini (SMR=13.9, CI=11.42, 14.31) (table 10, figure 14c).

Table 10. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death :"all cause CKD", "CKD attributed to DM" and "non DM related CKD" for males 15-49 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to DM	Non-DM related CKD
			SMR (95% CI)	
Central America	El Salvador	5.62 (5.35,5.89)	1.60 (1.44,1.75)	25.03 (23.67,26.39)
	Nicaragua	3.71 (2.34,4.56)	1.02 (1.01,2.30)	16.66 (16.07,17.69)
Oceania	Papua New Guinea	3.21 (2.96,4.43)	1.52 (1.40,1.64)	11.50 (11.41,11.96)
	Kiribati	3.21 (3.00,4.57)	1.57 (1.46.2.33)	11.14 (11.00,14.57)
	Solomon Islands	3.17 (2.96,4.59)	1.51 (1.50,2.34)	11.16 (10.67,12.38)
South Asia	India	1.26 (1.20,1.38)	0.41 (0.29,0.97)	5.33 (4.97,5.49)
Southern SSA	eSwatini	3.10 (3.00,3.59)	0.86 (0.76,0.97)	13.90 (11.42,14.31)
	Lesotho	2.67 (1.97,2.75)	0.69 (0.43,0.74)	12.24 (11.78,13.47)
Western SSA	Guinea-Bissau	2.30 (2.21,3.48)	0.38 (0.23.0.76)	11.54 (10.53,14.37)
	Sao Tome and Principe	1.84 (1.10,2.56)	0.28 (0.10,0.38)	9.36 (8.74,12.56)
Central SSA	Central African Republic	2.06 (2.00,4.31)	0.39 (0.22,0.99)	10.13 (10.00,11.21)
Eastern SSA	Eritrea	2.05 (2.01,5.69)	0.35 (0.21,0.44)	10.25 (9.74,14.33)
	Ghana	1.92 (1.10,3.47)	0.36 (0.10,0.49)	9.44 (9.00,14.22)

The high mortality rate ratios were largely seen in Africa, in Nigeria (8.29:1), Tanzania (7.65:1), and Niger (7.26:1) (figure 14d).

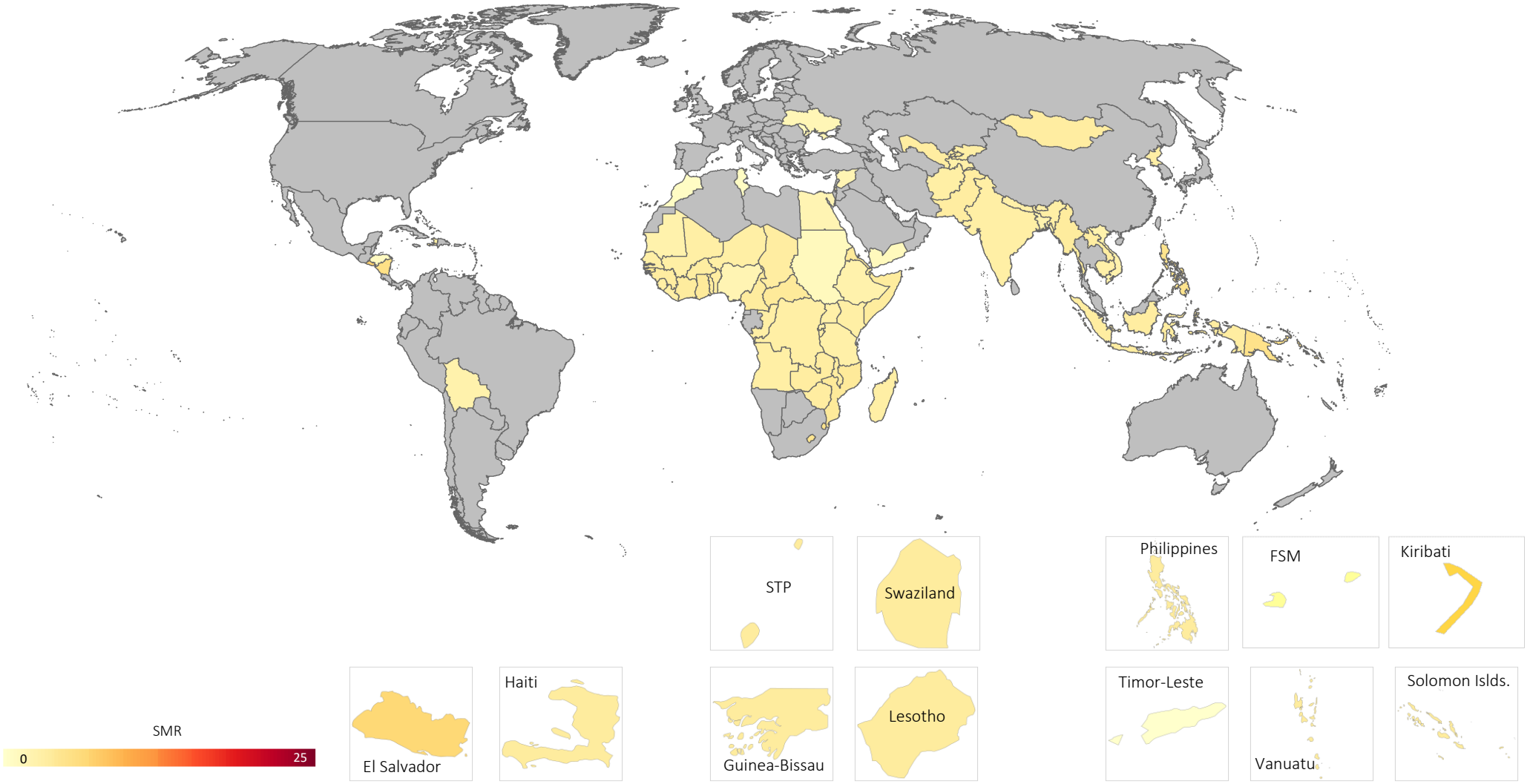


Figure 14a. Spatial distribution of SMR values for all-cause CKD for 15-49-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

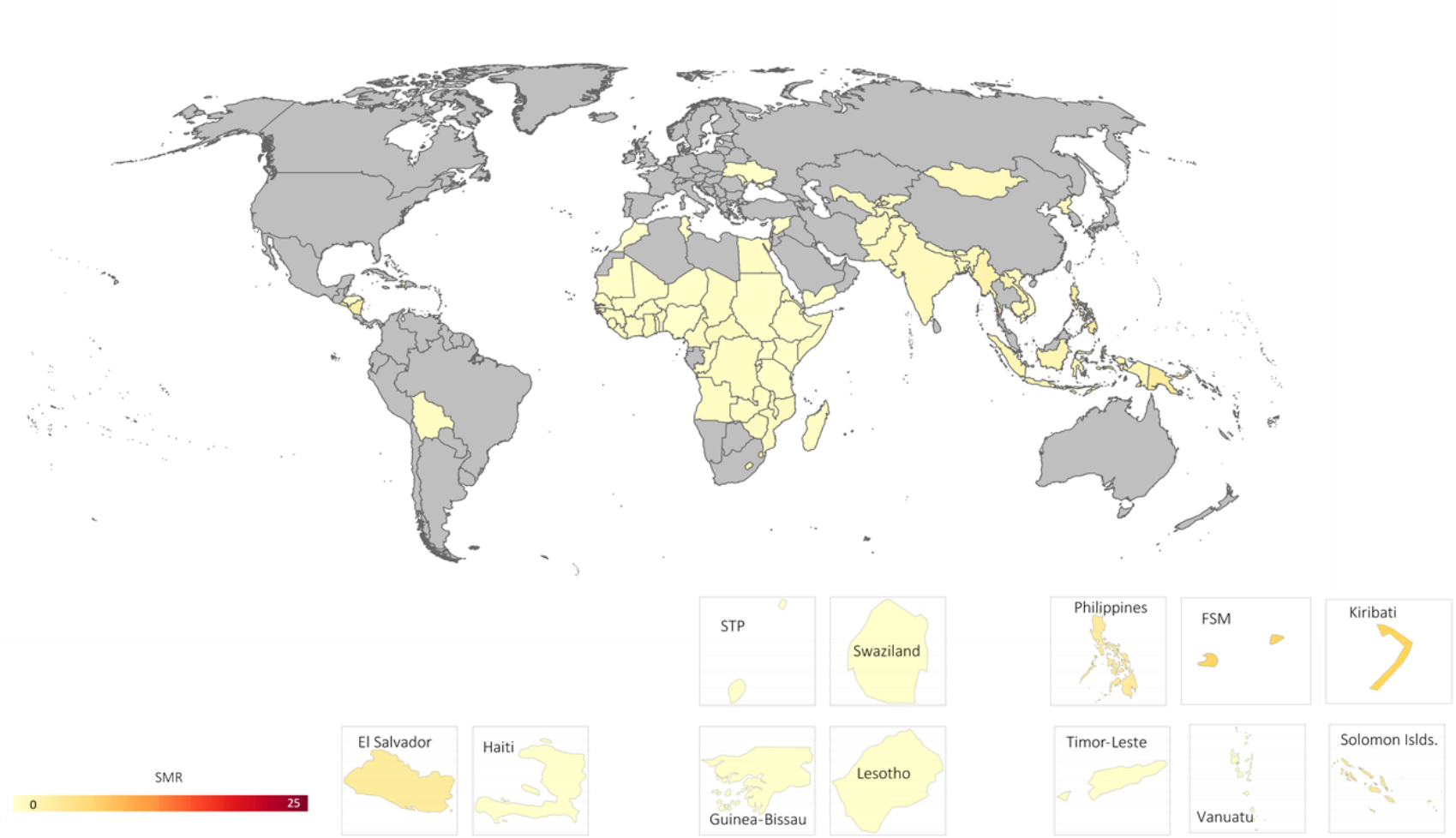


Figure 14b. Spatial distribution of SMR values for CKD attributed to diabetes for 15-59-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

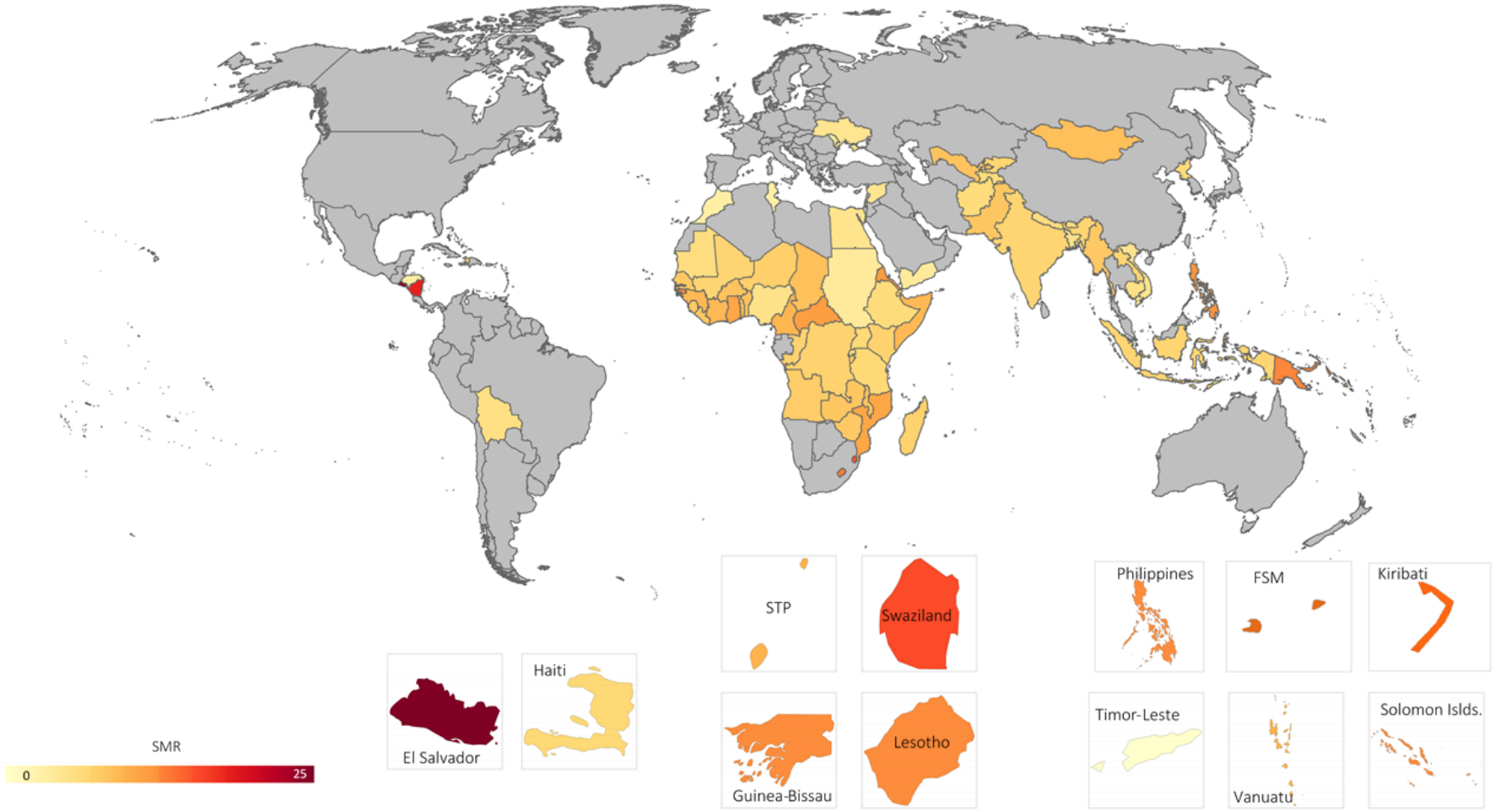


Figure 14c. Spatial distribution of SMR values for non-diabetes related CKD for 15-49-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

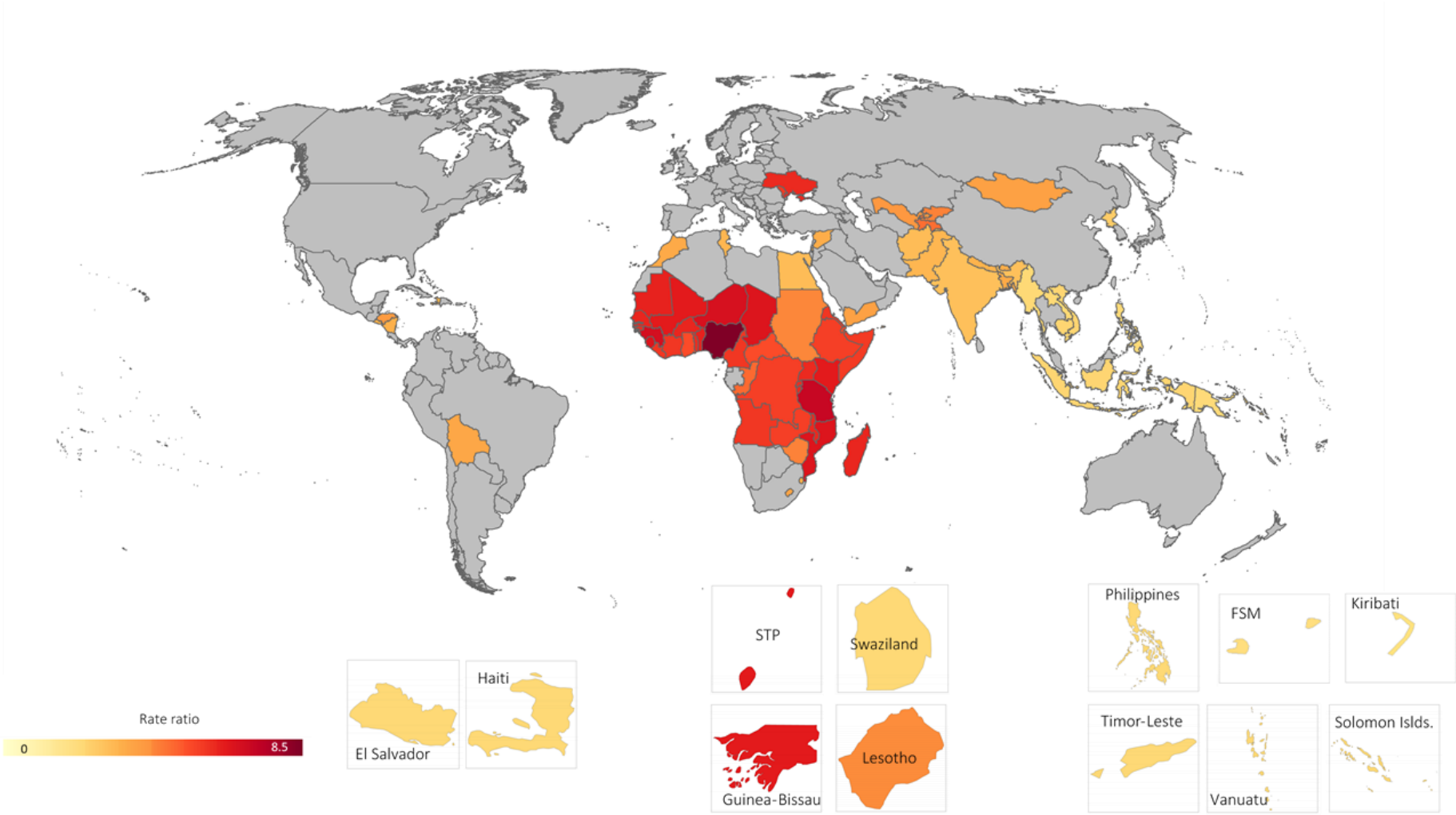


Figure 14d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 15-49-year-old in lower income countries

Males 50-69 years

Excess deaths were generally lower in the 50-69-year age group than for the 15-49-year age group, but broadly followed a similar geographical pattern.

The highest SMR for all-cause CKD was in El Salvador (SMR=4.34, CI=4.20, 4.48), which was significantly higher than the other regions (table 11, figure 15a). For DM CKD, excess mortality was lower than that for all cause CKD in both Central America and SSA. In Oceania however, the deaths were significantly higher in all countries (table 11, figure 15b). For non-DM CKD excess mortality was again significantly higher in El Salvador (SMR=4.35, CI=4.18, 4.52). In the Federated States of Micronesia (FSM), the excess mortality was significantly lower for (SMR=1.67 95%CI=1.11-1.92) than that observed DM CKD, denoting that a large proportion of mortality in this group is due to diabetes-related renal disease (table 11, figure 15c).

Table 11. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: "all cause CKD", "CKD attributed to DM" and "non-DM related CKD" for males 50-69 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to DM	Non-DM related CKD
			SMR (95% CI)	
Central America	El Salvador	4.34 (4.20,4.48)	4.31 (4.08, 4.56)	4.35 (4.18,4.52)
	Nicaragua	2.83 (1.97,3.44)	2.71 (2.37,3.56)	2.90 (1.57,3.45)
Southern SSA	eSwatini	2.29 (2.21,2.93)	2.10 (1.88,3.00)	2.39 (2.11,3.97)
	Lesotho	1.98 (1.59,2.33)	1.68 (1.49,2.22)	2.13 (1.11,2.35)
Western SSA	Guinea-Bissau	1.39 (1.20,3.44)	1.02 (1.00,3.55)	1.58 (1.27,2.35)
	Sao Tome and Principe	1.21 (1.07,2.31)	0.78 (0.44,0.98)	1.43 (1.42,3.66)
Central SSA	Central African Republic	1.32 (1.02,1.68)	1.06 (1.02,2.33)	1.47 (1.46,2.33)
Eastern SSA	Eritrea	1.12 (1.10,1.36)	0.94 (0.74,0.97)	1.21 (1.17,2.34)
Oceania	FSM	2.03 (2.00,2.41)	2.70 (2.11,3.44)	1.67 (1.11,1.92)
	Papua New Guinea	1.60 (1.44,1.97)	2.08 (1.97,2.05)	1.35 (1.21,1.46)
	Kiribati	1.43 (1.22,1.56)	2.02 (1.76,3.23)	1.21 (1.06,1.87)

	Solomon Islands	1.48 (1.47,2.59)	1.96 (1.46,3.57)	1.23 (1.27,3.32)
South Asia	India	0.96 (0.89,0.97)	1.40 (1.38,1.59)	0.73 (0.68,0.89)

The higher rate ratios for non-DM CKD: DM-CKD were seen in the Ukraine (4.19:1) and Nigeria (3.77:1) (figure 15d).

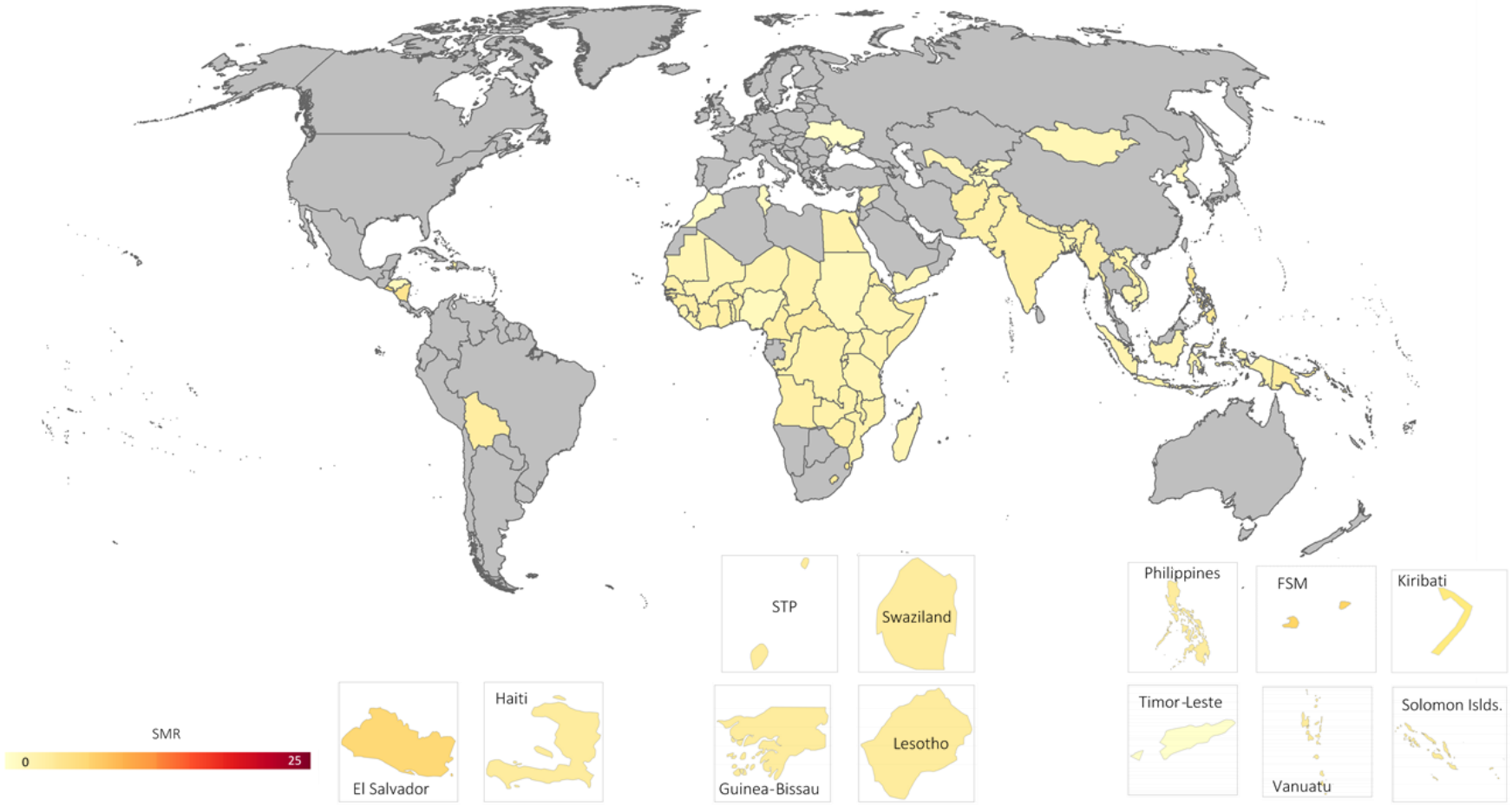


Figure 15a. Spatial distribution of SMR values for all-cause CKD for 50-69-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

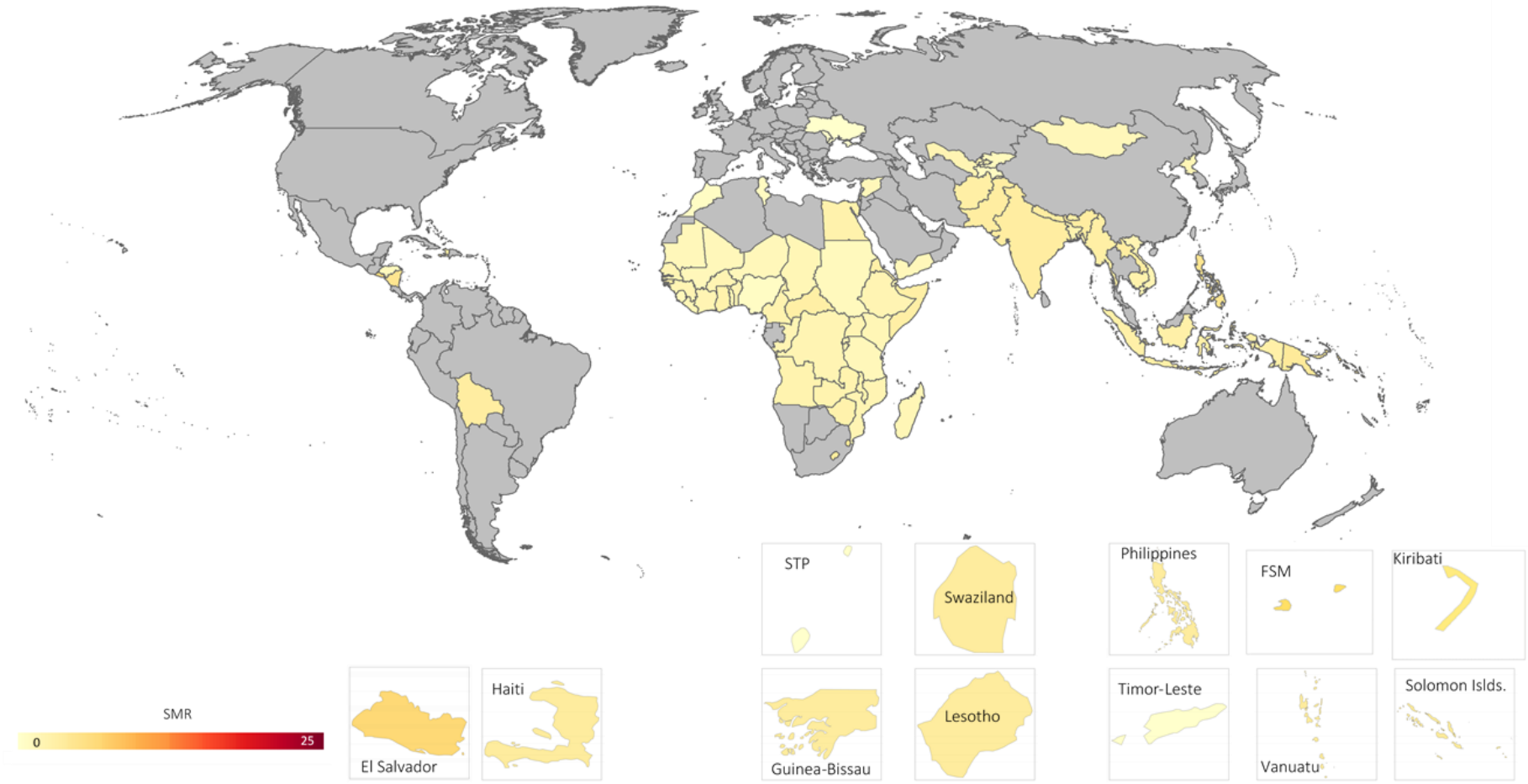


Figure 15b. Spatial distribution of SMR values for CKD attributed to diabetes CKD for 50-69-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

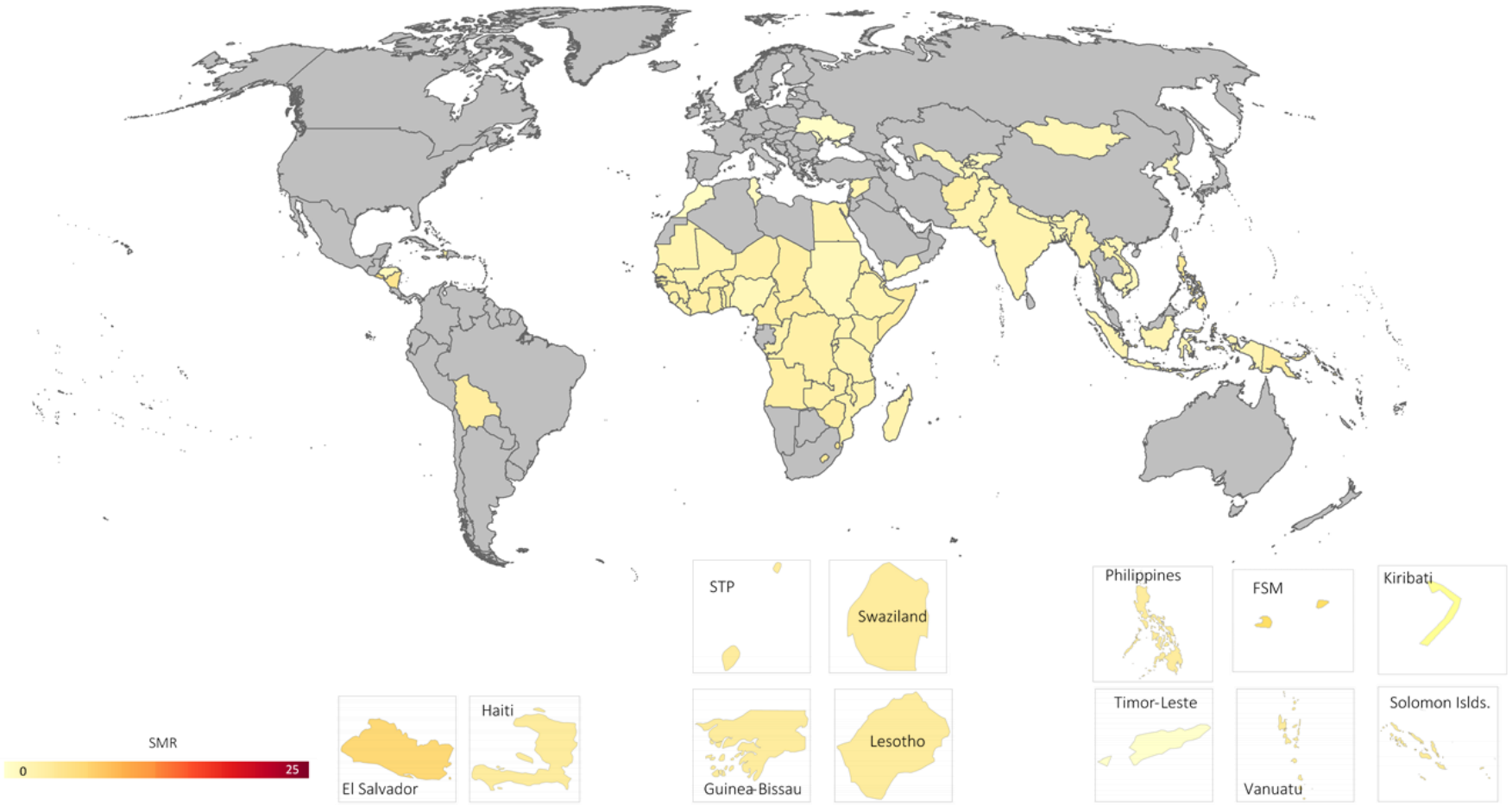


Figure 15c. Spatial distribution of SMR values for non-diabetes related CKD for 50-69-year-old males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

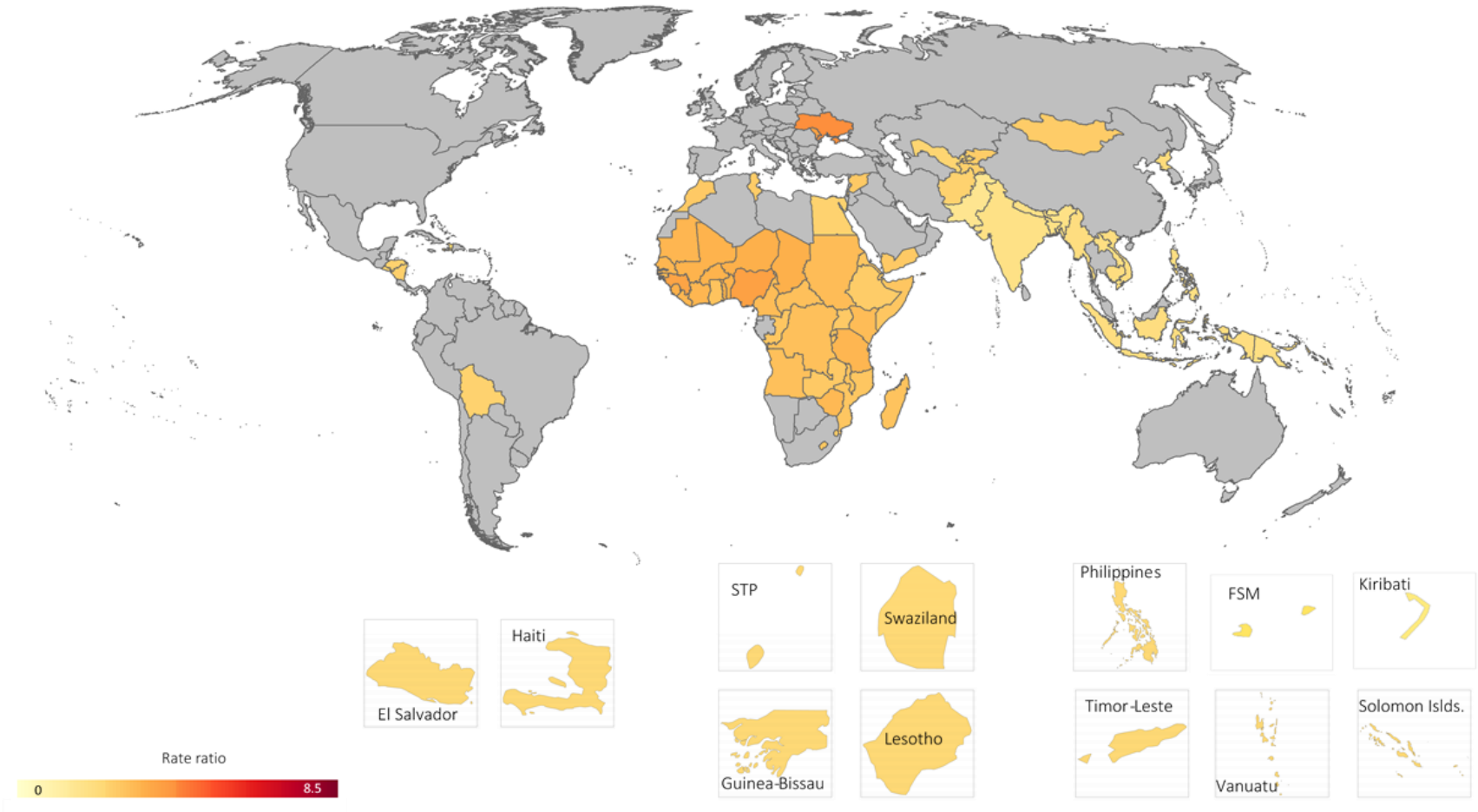


Figure 15d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 50-69-year-old in lower income countries

Males 70+ years

For this age group excess mortality was seen in fewer countries, and excess mortality was markedly lower in comparison to younger age categories.

For both all cause CKD and diabetic CKD excess deaths were seen only in El Salvador (SMR=1.79, 95%CI=1.74, 1.84, and SMR=1.80, 95%CI=1.70, 1.91) and the Federated States of Micronesia (SMR=1.18, 95%CI=0.64, 1.72, and SMR=1.72, 95%CI=1.50, 2.93). In both cases, the diabetes-attributed CKD deaths were the highest of all causes of death (table 12, figure 16a and figure 16b respectively). For non-diabetic CKD group, excess deaths were only seen in El Salvador (SMR=1.78, 95%=1.73, 1.84) (table 12, figure 16c).

Table 12. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death "all cause CKD", "CKD attributed to diabetes" and "non diabetes related CKD" for males 70+ years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to diabetes	Non-diabetes related CKD
			SMR (95% CI)	
Central America	El Salvador	1.79 (1.74,1.84)	1.80 (1.70,1.91)	1.78 (1.73,1.84)
Oceania	FSM	1.18 (0.64,1.72)	1.72 (1.50,2.93)	0.96 (0.33,0.97)

The higher rate ratios were seen in Nigeria (5.21:1), Guinea (4.78:1) and Moldova (4.75:1) (figure 16d).

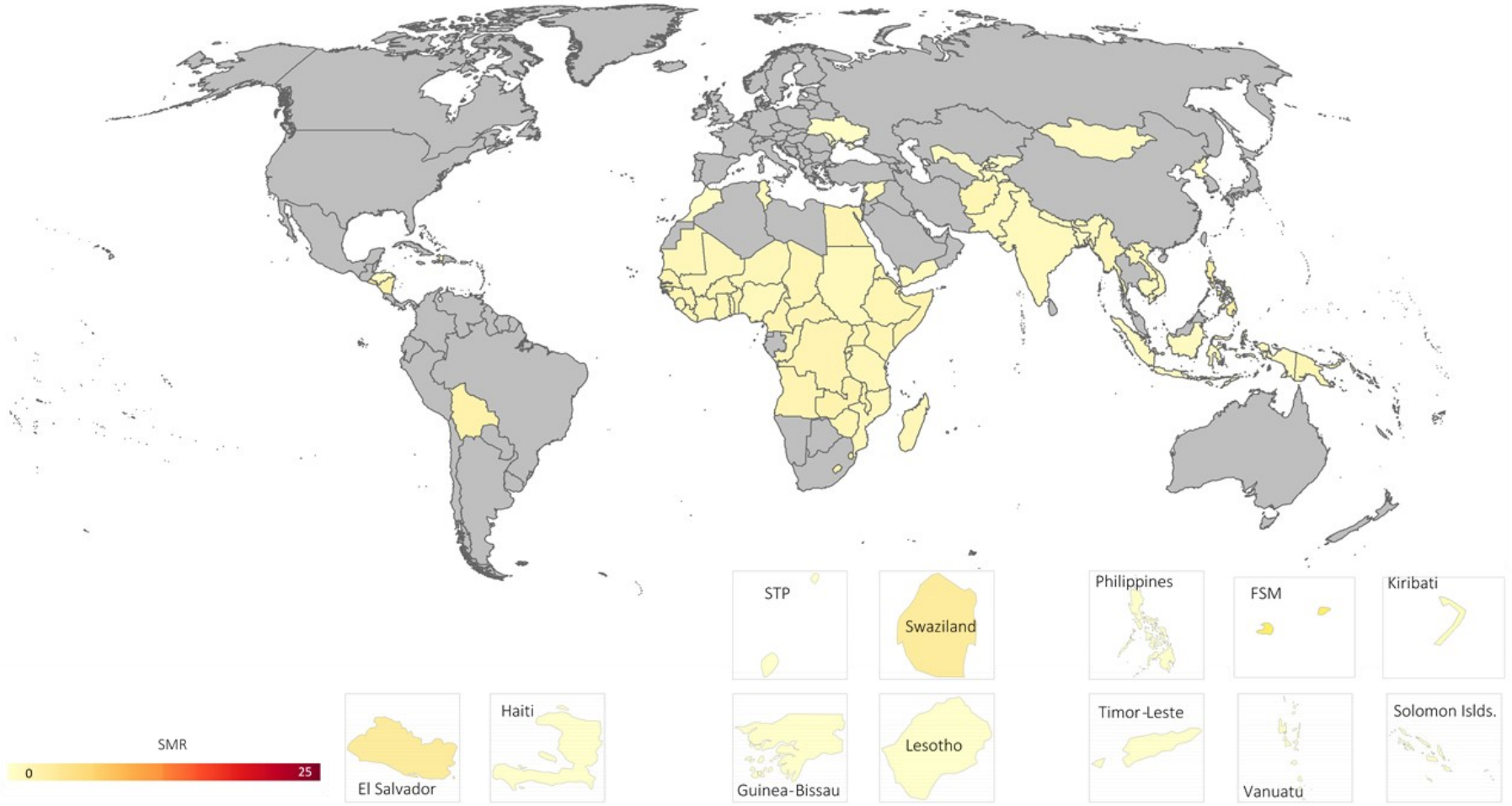


Figure 16a. Spatial distribution of SMR values for all-cause CKD for 70+ year males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

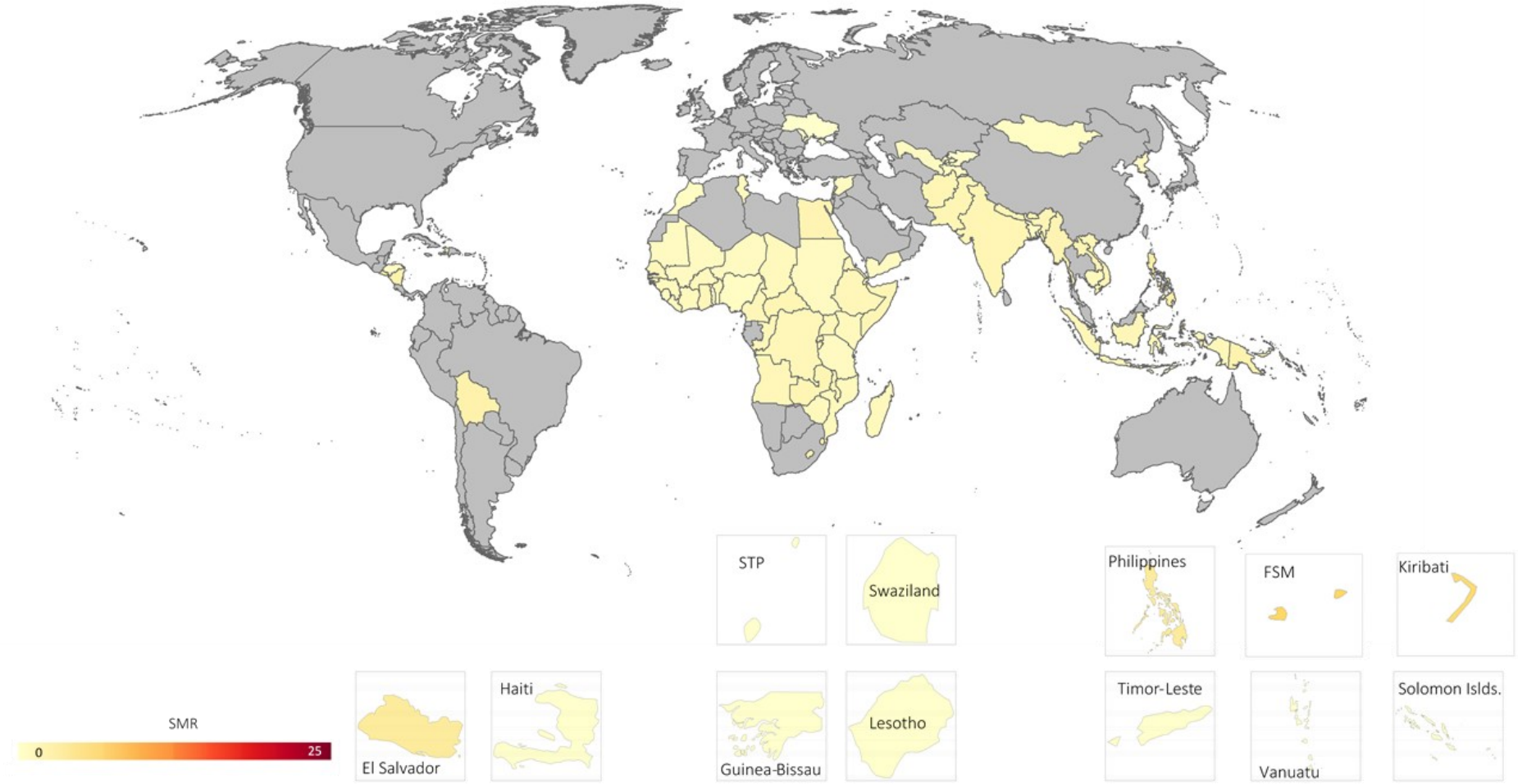


Figure 16b. Spatial distribution of SMR values for CKD attributed to diabetes for 70+ year males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

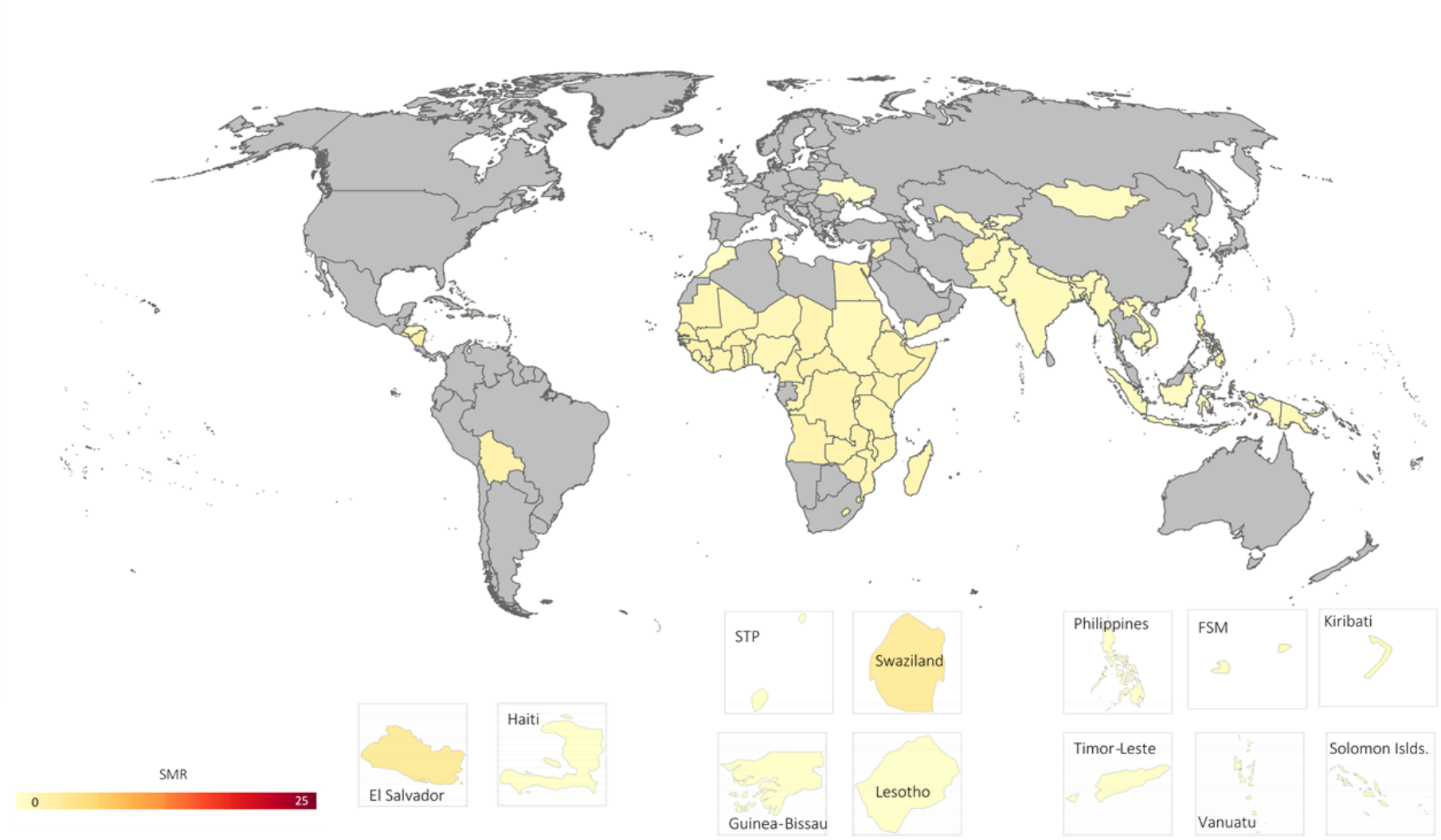


Figure 16c. Spatial distribution of SMR values for non-diabetes related CKD for 70+ year males in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

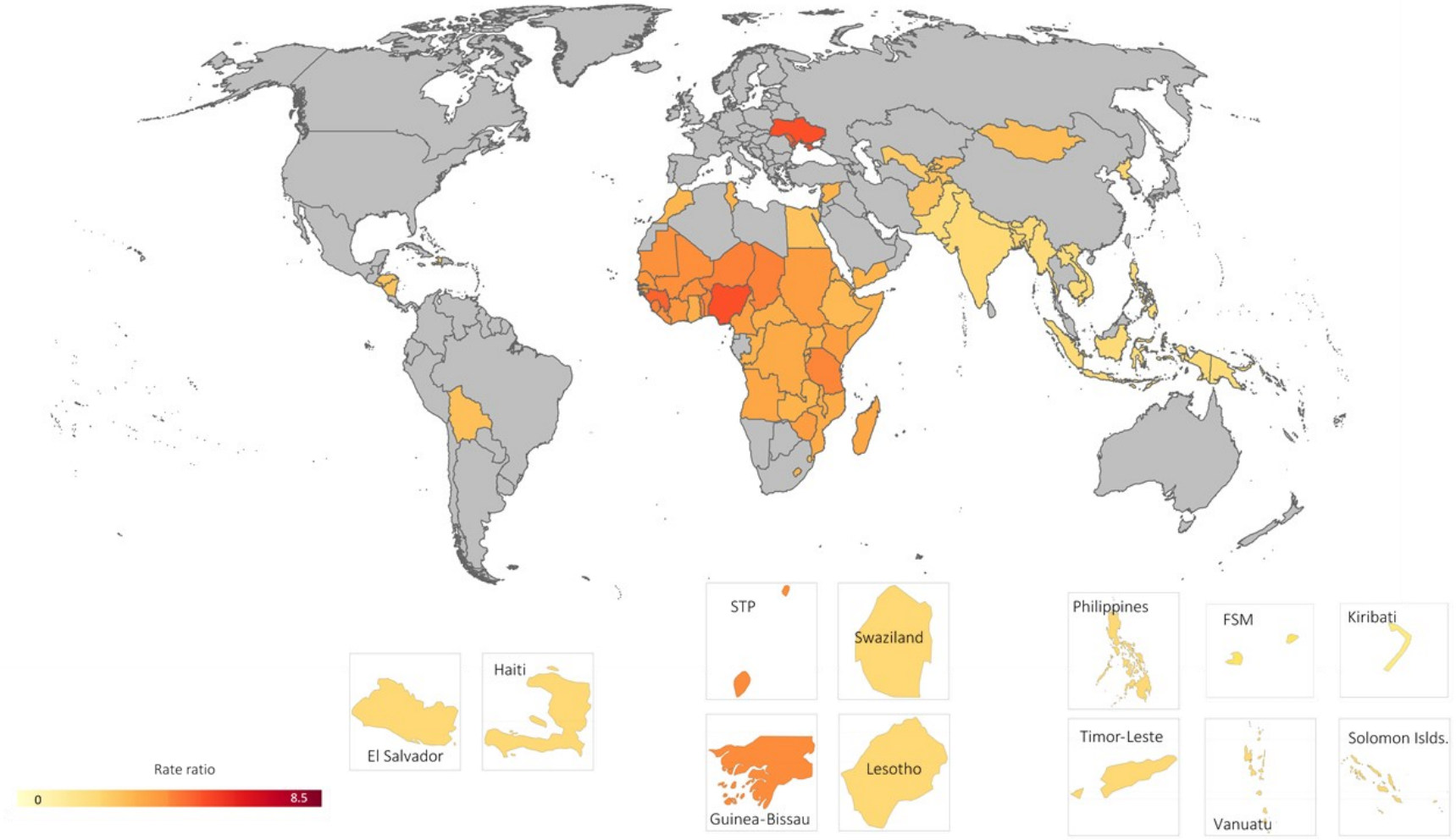


Figure 16d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for males 70+ years in lower income countries

Females

15-49 years

For females in the lower-middle income countries, excess deaths were observed across were Oceania, Africa Central America and South Asia. For all-cause CKD the Solomon Islands had the highest overall excess mortality (SMR=5.62, 95%CI=4.91,6.33), which was higher than the other detected regions. There was also high excess mortality detected in south Asia, the highest of which was in Pakistan (SMR=3.01, 95%CI=2.97,3.05). In Western Sub-Saharan Africa excess deaths were also detected and were higher than those in Central America (table 13, figure 17a). For DM CKD excess deaths were the lowest of all mortality categories. The Solomon Islands again had higher mortality than the other detected regions (3.95, 95%CI=2.87,5.03). No excess deaths were observed in SSA (table 13, figure 17b). The excess deaths for non-DM CKD were the highest observed for this age group and were markedly higher than DM CKD in every region. The Solomon Islands and Pakistan again had the highest SMR (SMR=7.79 95%CI=6.07,9.52 and SMR=5.62 95%CI=5.54,5.70 respectively) (table 13, figure 17c).

Table 13. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: "all cause CKD", "CKD attributed to diabetes" and "non diabetes related CKD" for females 15-49 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to diabetes	Non-diabetes related CKD
			SMR (95% CI)	
Oceania	Solomon Islands	5.62 (4.91,6.33)	3.95 (2.87,5.03)	7.79 (6.07,9.52)
	FSM	4.58 (4.51,5.76)	3.31 (2.22,4.55)	6.22 (6.11,8.44)
	Papua New Guinea	4.06 (4.00,4.97)	2.84 (2.61,3.12)	5.65 (5.11,5.55)
South Asia	Pakistan	3.01 (2.97,3.05)	1.10 (1.08,3.00)	5.62 (5.54,5.70)
	Myanmar	2.59 (2.34,2.99)	1.75 (1.69,1.87)	3.67 (3.19,4.22)
	Nepal	1.60 (1.48,1.77)	0.60 (0.42,0.57)	2.89 (1.99,3.00)
	India	1.22 (1.19,1.38)	0.45 (0.38,0.91)	2.22 (1.96,2.38)

Central America	El Salvador	2.03 (2.00,2.41)	1.13 (1.11,3.44)	4.05 (3.11,4.95)
	Nicaragua	1.80 (1.32,2.00)	0.84 (0.32,0.87)	3.07 (2.47,4.02)
Western SSA	Guinea-Bissau	2.40 (2.33,3.11)	0.73 (0.02,0.78)	4.02 (4.00,4.51)
	Sao Tome and Principe	3.72 (3.31,4.66)	1.08 (1.01,1.24)	7.16 (6.88,8.91)
Central SSA	Congo	1.91 (1.32,2.77)	0.80 (0.32,0.89)	3.35 (3.27,4.21)
Southern SSA	Lesotho	1.31 (1.13-3.22)	0.60 (0.44,0.96)	2.23 (2.20,4.55)

The rate ratio maps showed the highest values in Rwanda (5.57:1), Mongolia (5.01:1) and Nigeria (4.98:1) (figure 17d).

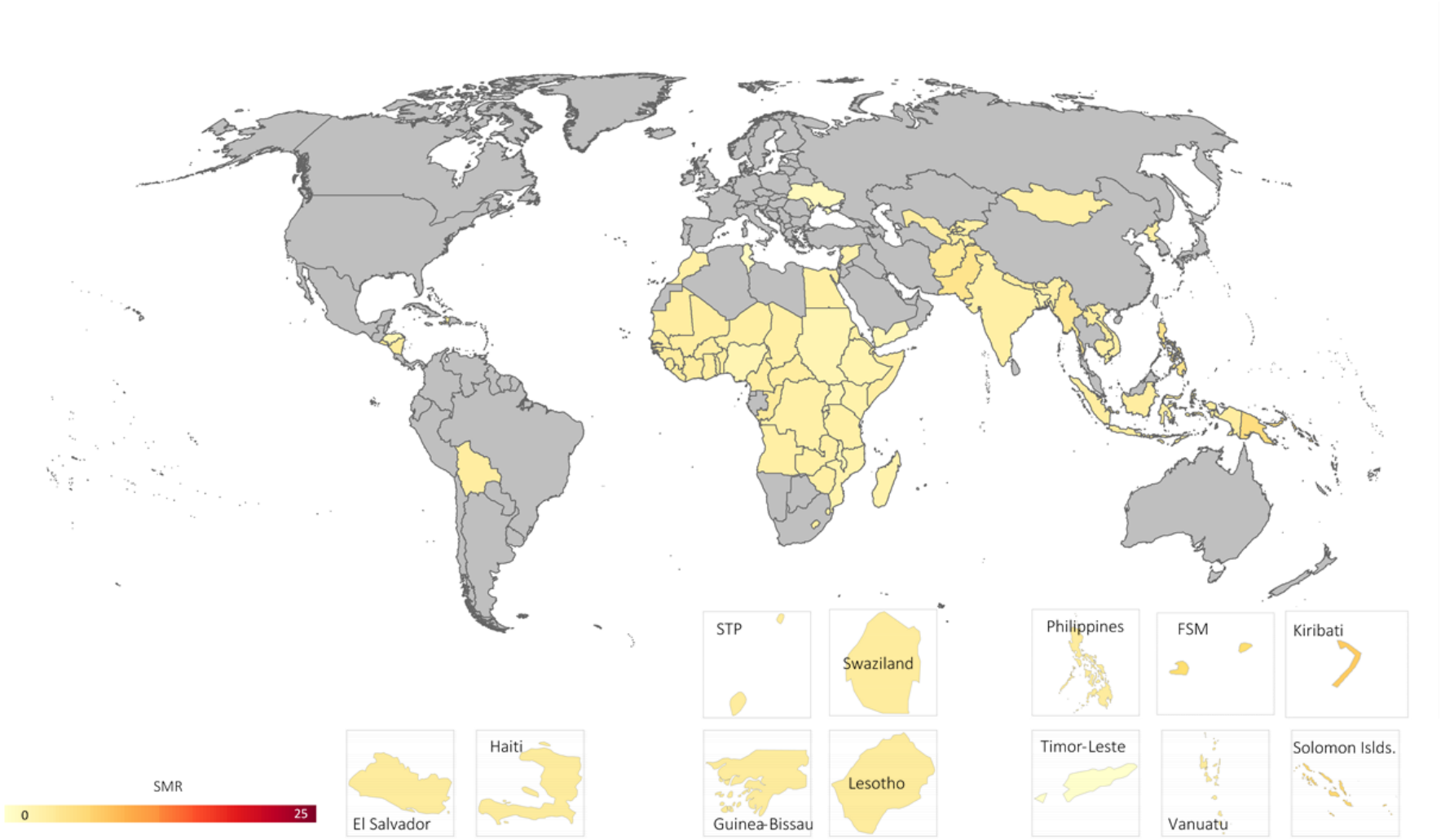


Figure 17a. Spatial distribution of SMR values for all-cause CKD for 15-49-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

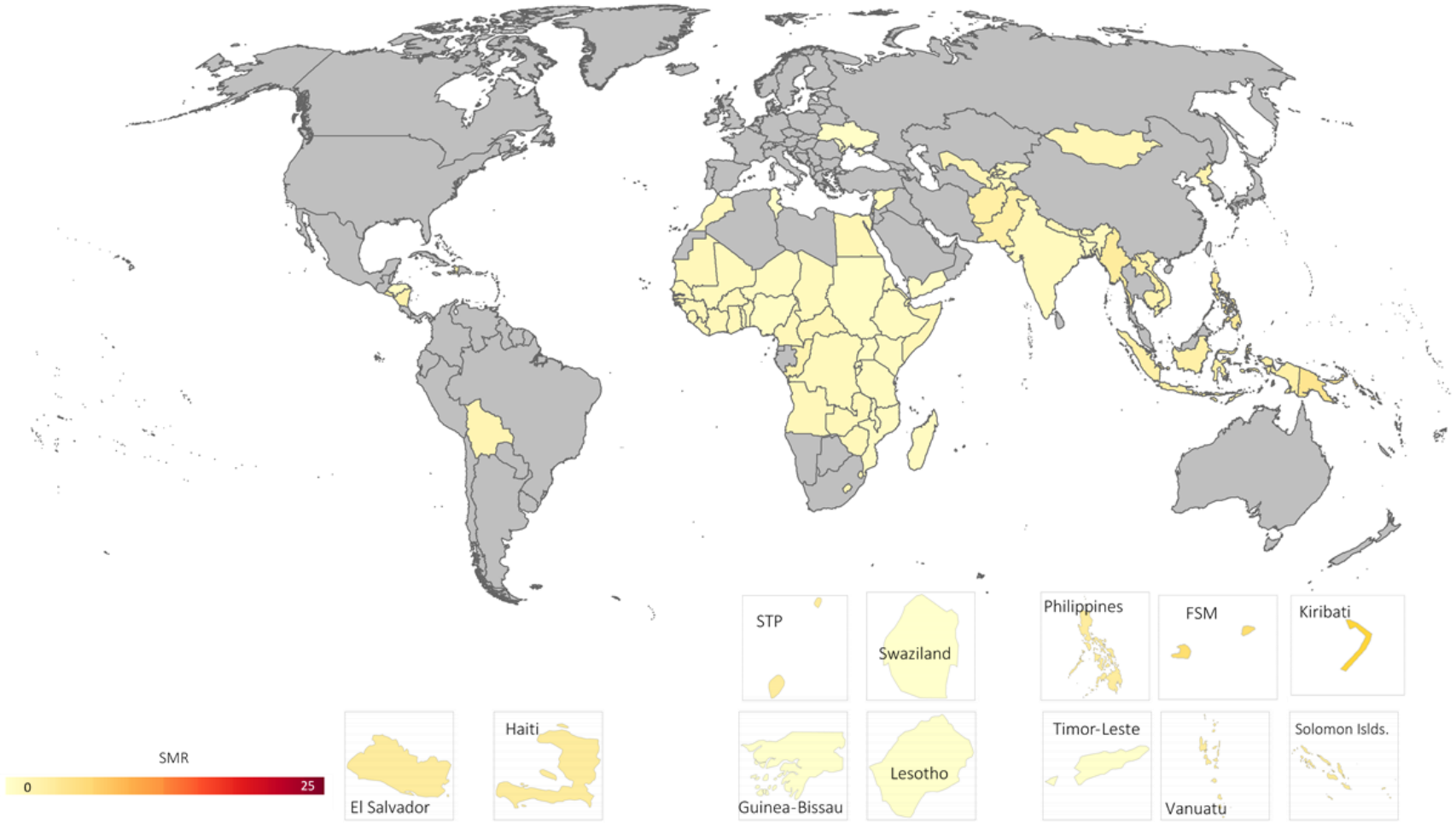


Figure 17b. Spatial distribution of SMR values for CKD attributed to diabetes for 15-49-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

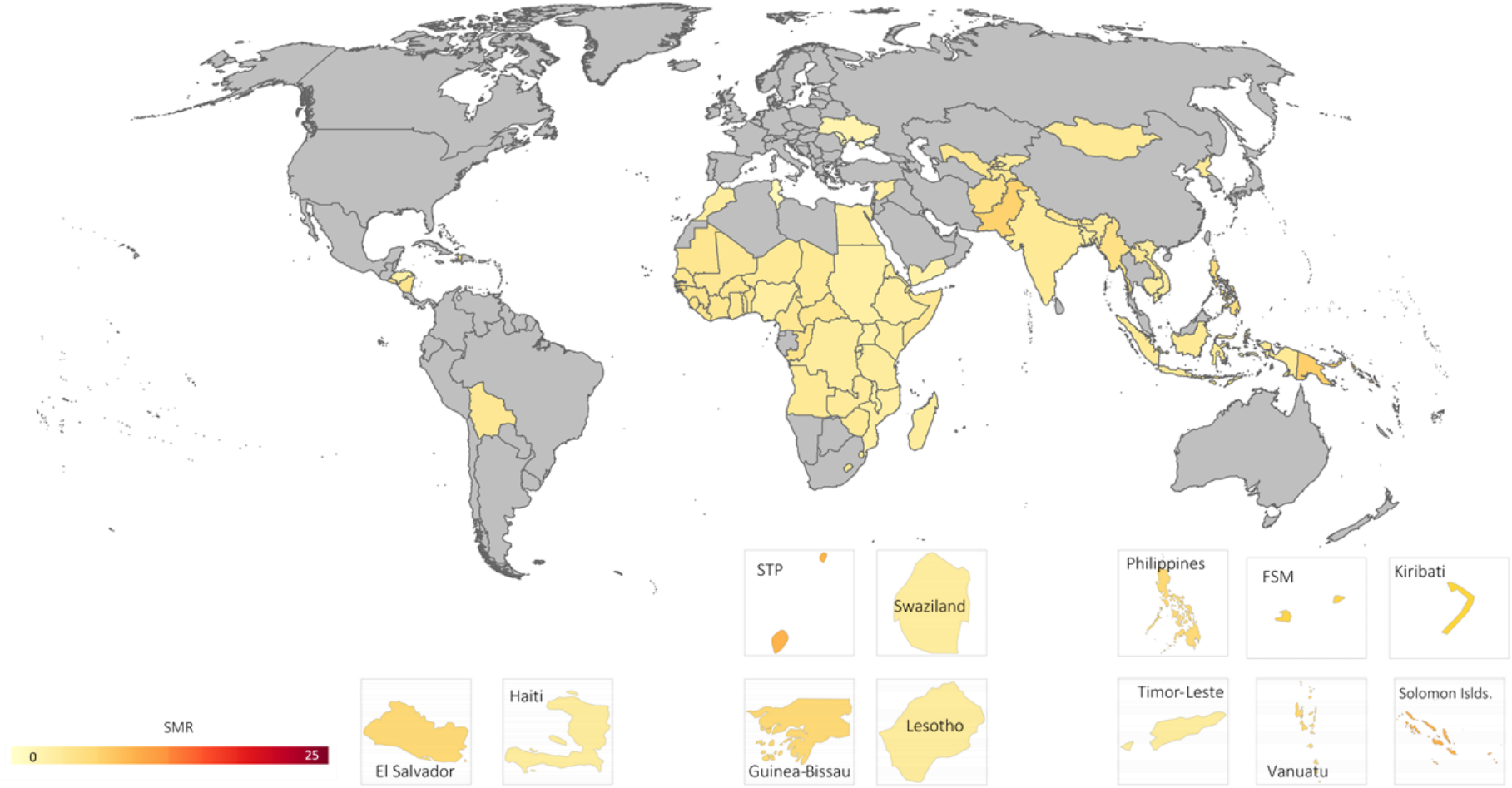


Figure 17c. Spatial distribution of SMR values for non-diabetes related CKD for 15-49-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

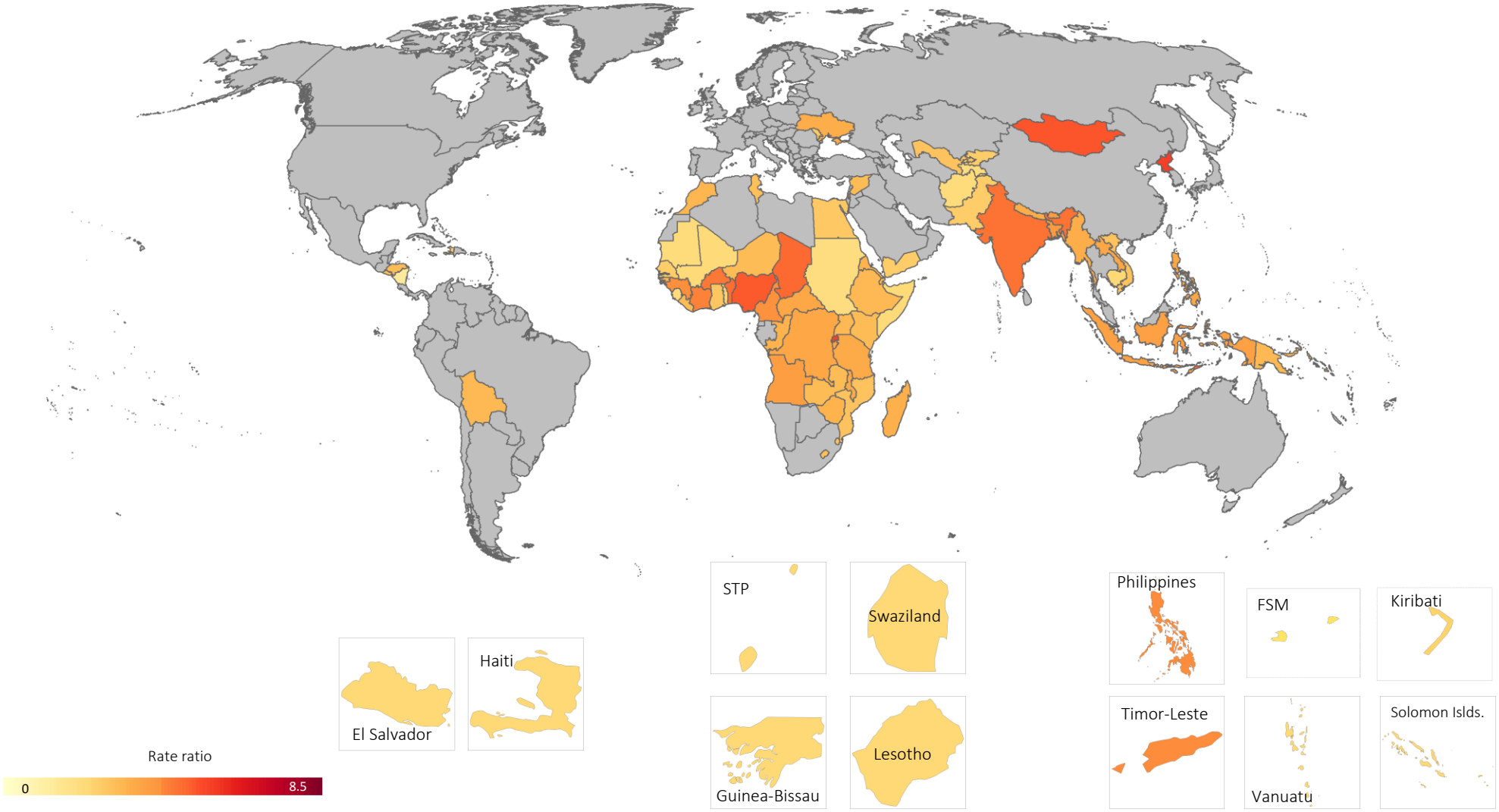


Figure 17d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for 15-49-year females in lower income countries

Females 50-69 years

For all cause CKD the same geographical pattern was observed as 15-49 year old females, with the highest excess deaths seen largely across Oceania. There were fewer excess deaths in Central American countries, however the highest remained in El Salvador (SMR=2.79, 95%CI=1.84,3.22). (table 14, figure 18a). Excess deaths for DM CKD were higher highest in Oceania and South Asia, the highest of which was in the Federated States of Micronesia (SMR=4.56, 95%CI=2.36,6.75). (table 14, figure 18b). For non-DM CKD, the SMR values were lower than those for "all cause CKD", and largely followed the same geographical distribution as before, with the Federated States of Micronesia having the highest excess deaths (SMR=2.72, 95%CI=1.96,2.58), followed by El Salvador (SMR=2.62, 95%CI=1.60,2.80) and Sao Tome and Principe (SMR=2.33, 95%CI=2.30,2.49) (table 14, figure 18c).

Table 14. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: "all cause CKD", "CKD attributed to diabetes" and "non diabetes related CKD" for females 50-69 years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to diabetes	Non-diabetes related CKD
			SMR (95% CI)	
Oceania	Federated States of Micronesia	3.36 (2.25,4.48)	4.56 (2.36,6.75)	2.72 (1.96,2.58)
	Solomon Islands	2.61 (2.54,3.91)	3.52 (2.34,4.21)	2.12 (1.76,3.11)
	Papua New Guinea	2.42 (1.97,2.67)	3.19 (3.11,4.12)	2.01 (1.56,3.27)
	Kiribati	2.21 (2.04,3.19)	3.05 (2.98,3.44)	1.76 (1.47,2.18)
South Asia	Pakistan	2.04 (1.92,2.10)	2.83 (2.56,3.11)	1.61 (1.49,3.27)
	Myanmar	1.46 (1.28,1.55)	1.88 (1.28,1.92)	1.23 (1.20,1.34)
	Nepal	1.39 (1.21,1.89)	1.78 (1.43,2.22)	1.18 (1.02,2.31)
	India	1.16 (1.02,2.54)	1.47 (1.22,2.37)	0.79 (0.74,0.99)
Central America	El Salvador	2.79 (1.84,3.22)	3.11 (2.89, 3.33)	2.62 (1.60,2.80)
	Nicaragua	2.41 (1.98,2.66)	2.62 (2.12,3.68)	2.29 (1.95,3.01)
Western SSA	Sao Tome and Principe	2.05 (2.01,2.98)	1.55 (1.41,2.02)	2.33 (2.30,2.49)

Eastern SSA	Somalia	1.64 (1.57,2.11)	1.72 (1.61,3.44)	1.60 (1.37,3.55)
Southern SSA	Lesotho	1.30 (1.03,2.24)	1.31 (1.22,1.48)	1.29 (1.26,2.37)

The highest rate ratios were again seen in countries which had the lower SMR values for each CKD mortality cause (Ukraine 3.12:1, Moldova 2.97:1, and Nigeria 2.95:1 (figure 18d).

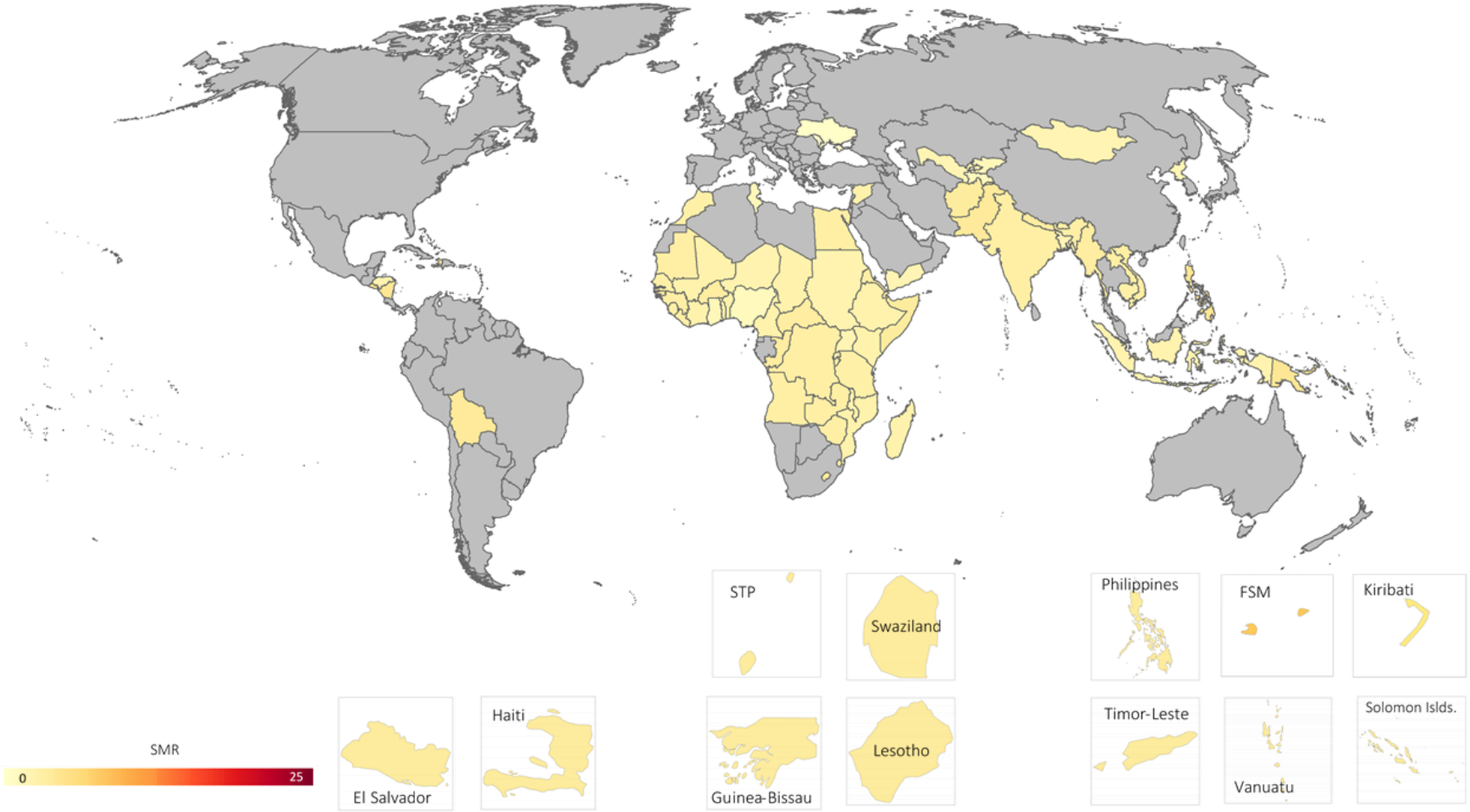


Figure 18a. Spatial distribution of SMR values for all-cause CKD for 50-69-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

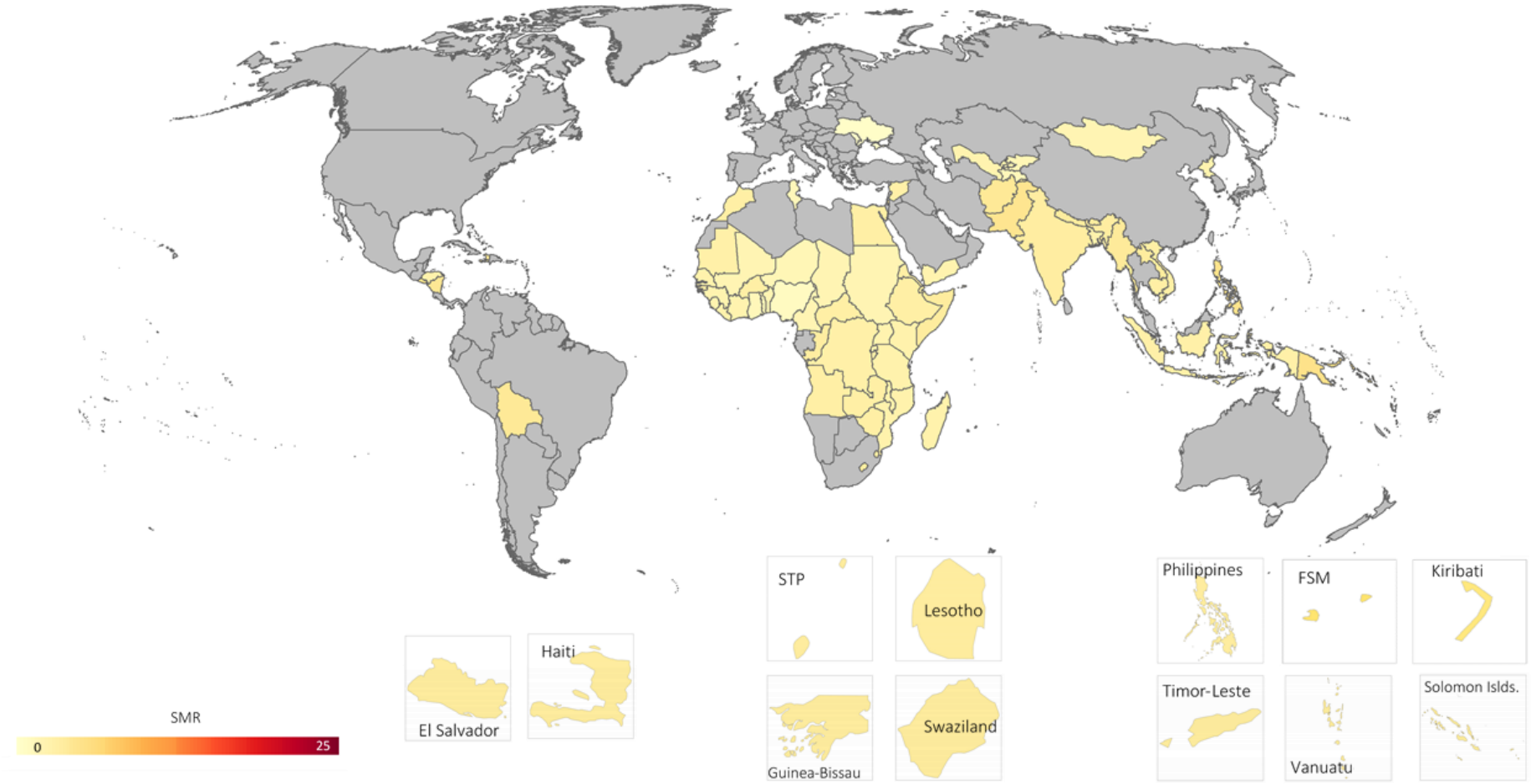


Figure 18b. Spatial distribution of SMR values for CKD attributed to diabetes for 50-69-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

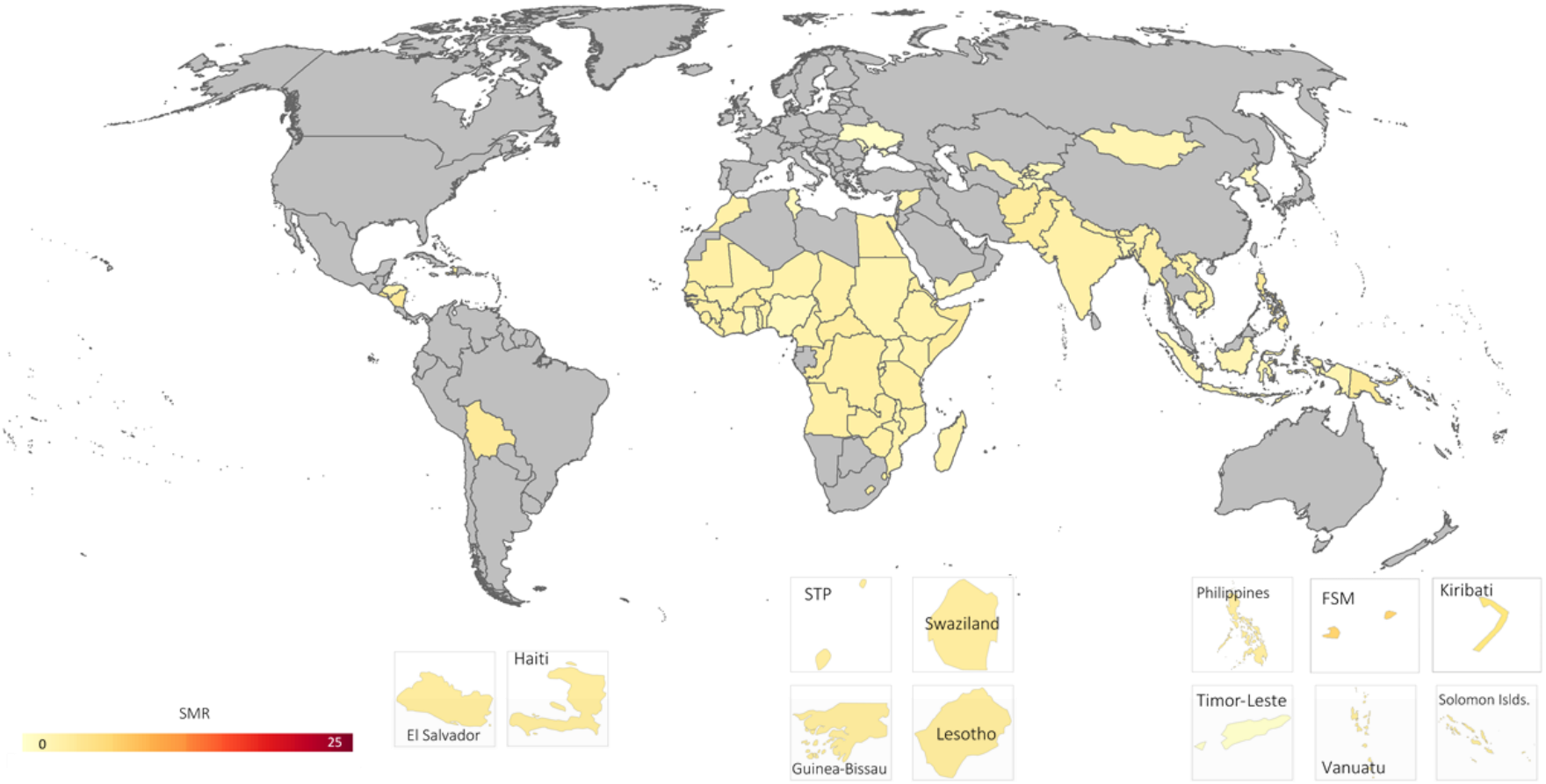


Figure 18c. Spatial distribution of SMR values for non-diabetes-related CKD for 50-69-year-old females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

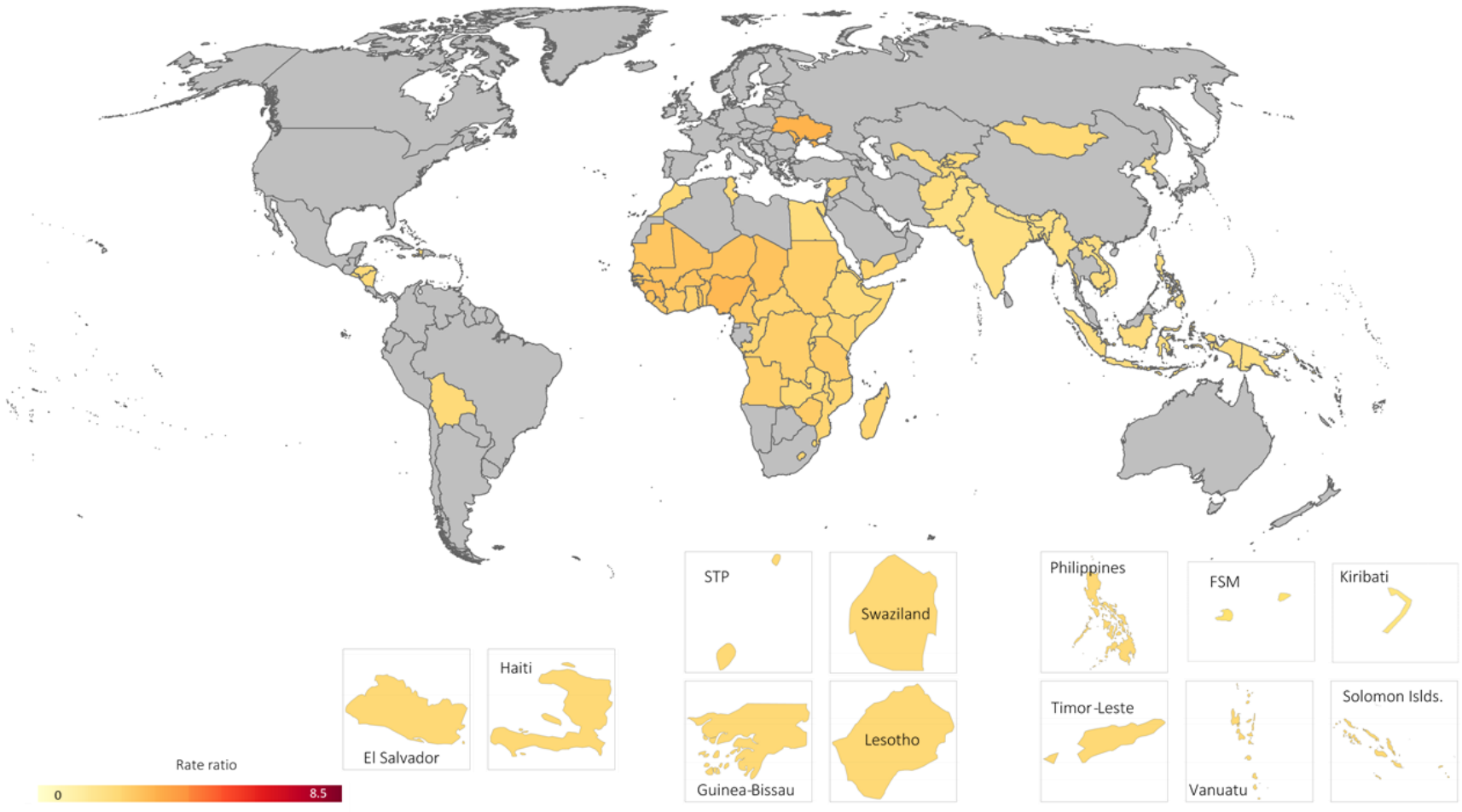


Figure 18d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related CKD for females 50-69 years in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

Females 70+ years

Like all other age and sex groups in the lower income category, the SMR values were the lowest for the 70+ age category. For all-cause CKD excess mortality was higher in the Federated States of Micronesia (SMR=2.10, CI=1.27, 2.93), which was markedly higher than those observed in El Salvador (SMR=1.81, CI=1.75, 1.87) and Sao Tome and Principe (SMR=1.66 95%CI=1.60-2.19) (table 15, figure 19a). Excess deaths for DM CKD were the highest across Oceania and Central America, however, none were detected in Africa (table 15, figure 19b).

For non-DM CKD the same ratio of excess mortality was observed in the Federated States of Micronesia (SMR=1.75, CI=1.68, 1.83) and El Salvador (SMR=1.75 95%CI=1.67-1.91) (table 15, figure 19c).

Table 15. Countries with the highest overall excess mortality (shown as SMR values) for each cause of death: "all cause CKD", "CKD attributed to diabetes" and "non diabetes related CKD" for females 70+ years. An SMR value <1 indicates a lower risk of mortality with respect to the global benchmark, and an SMR higher than 1 indicates a higher risk with respect to the global benchmark

Region	Country	All cause CKD	CKD attributed to diabetes	Non-diabetes related CKD
			SMR (95% CI)	
Oceania	Federated States of Micronesia	2.10 (1.27,2.93)	2.88 (1.15,4.61)	1.75 (1.68,1.83)
Central America	El Salvador	1.81 (1.75,1.87)	1.94 (1.82,2.06)	1.75 (1.67,1.91)
	Nicaragua	1.44 (1.34,1.97)	1.50 (1.34,2.44)	1.41 (1.39,1.78)
Western SSA	Sao Tome and Principe	1.66 (1.60,2.19)	1.05 (1.02,2.34)	1.94 (1.81,2.00)
South Asia	Pakistan	1.26 (1.12,3.45)	1.68 (1.51,2.54)	1.07 (1.02,1.35)

Like the other age groups and regions in these analyses, the higher rate ratios were seen in Nigeria (4.21:1), Sao Tome (4.08:1) and the Ukraine (3.87:1) (figure 19d).

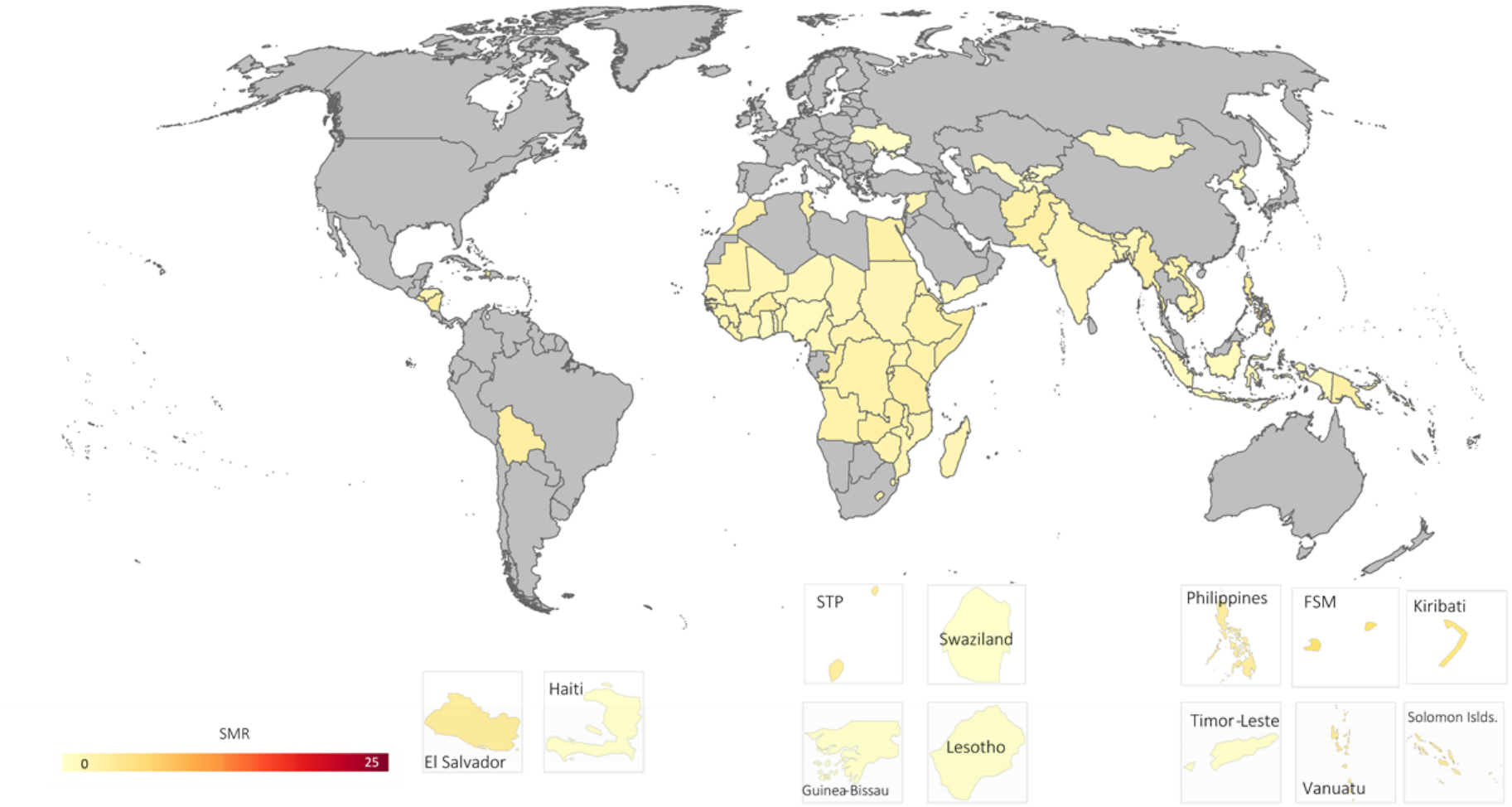


Figure 19a. Spatial distribution of SMR values for all-cause CKD for 70+ year females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

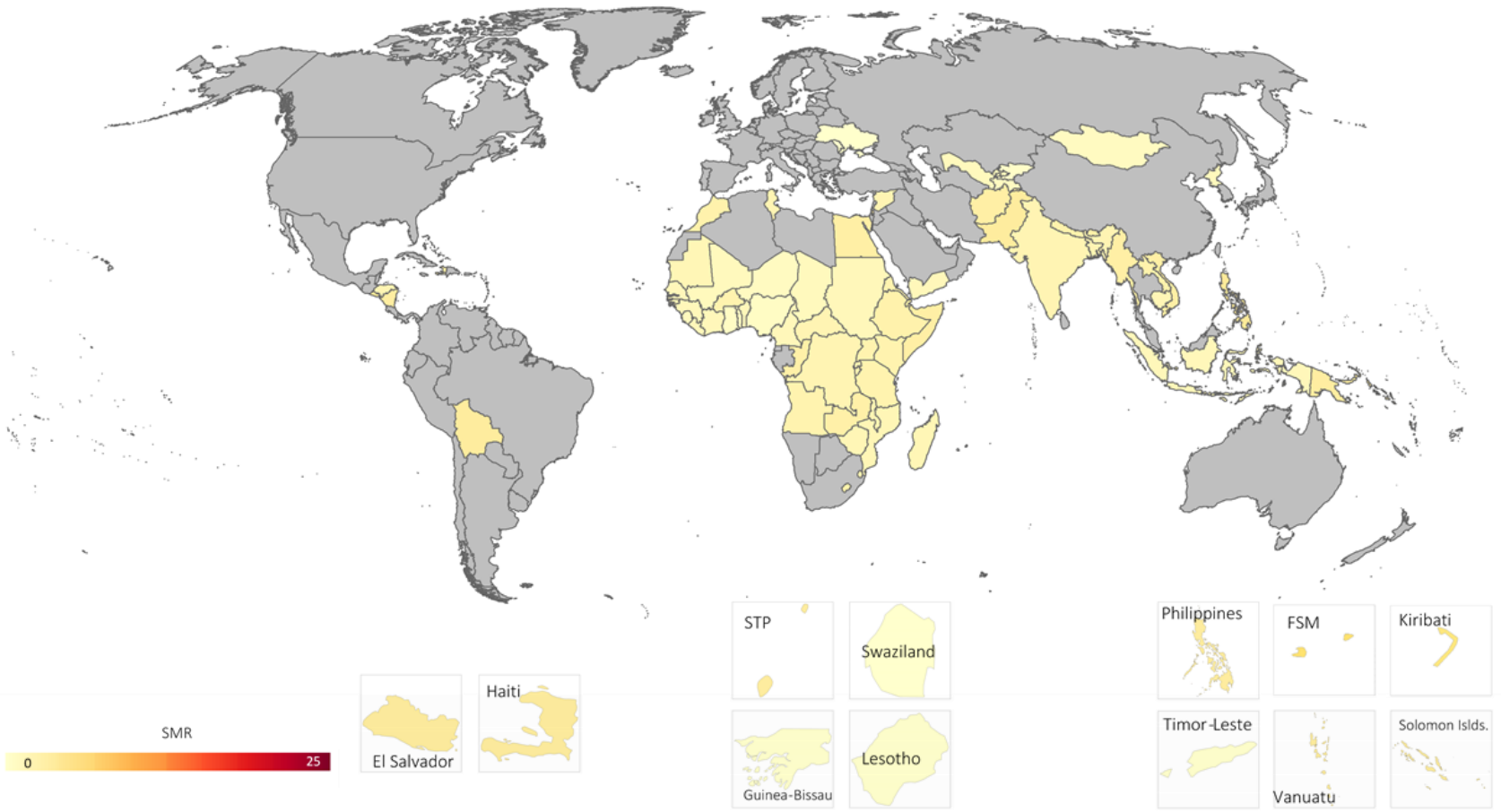


Figure 19b. Spatial distribution of SMR values for CKD attributed to diabetes for 70+ year females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

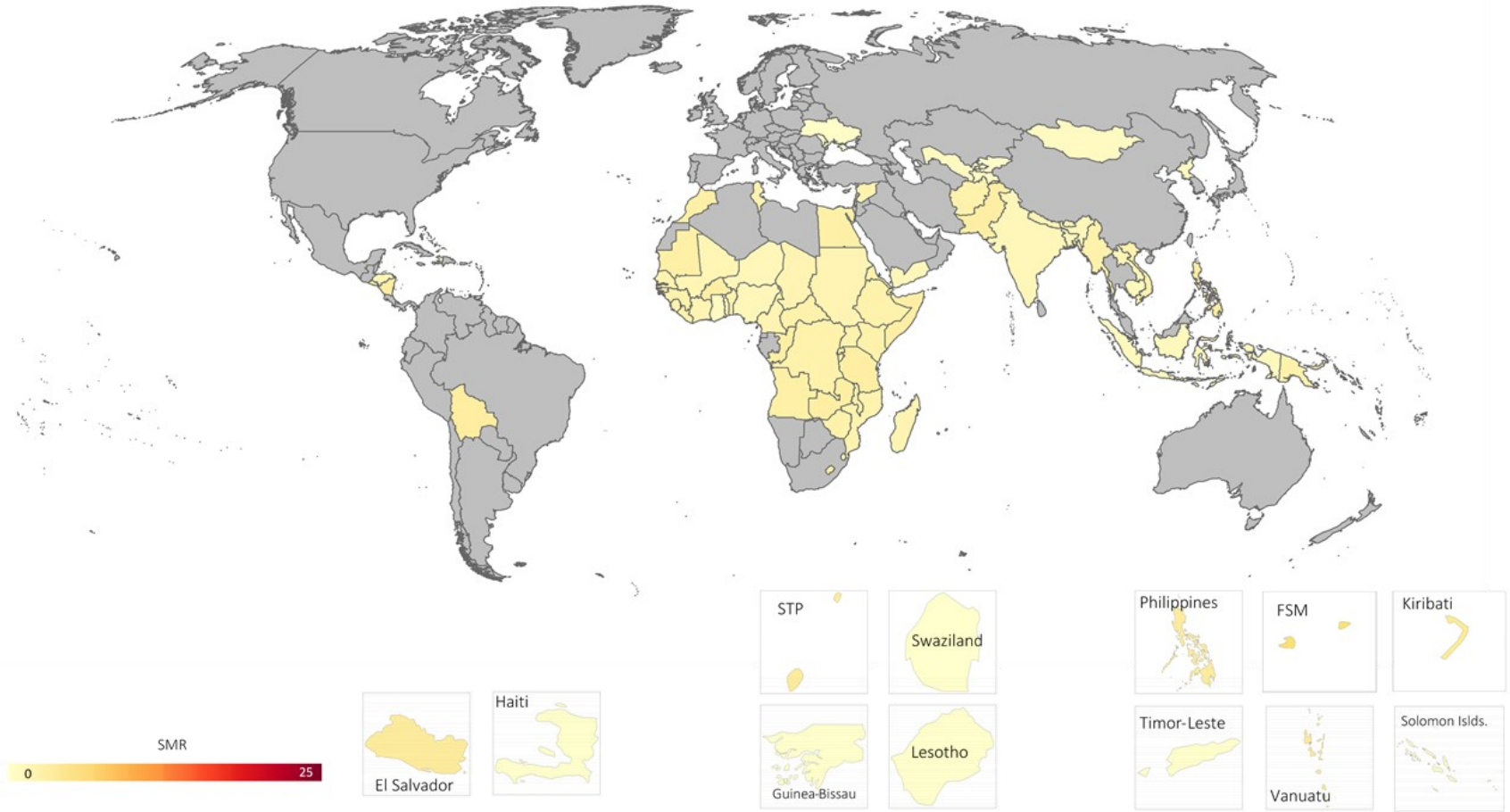


Figure 19c. Spatial distribution of SMR values for non-diabetes-related CKD for 70+ year females in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

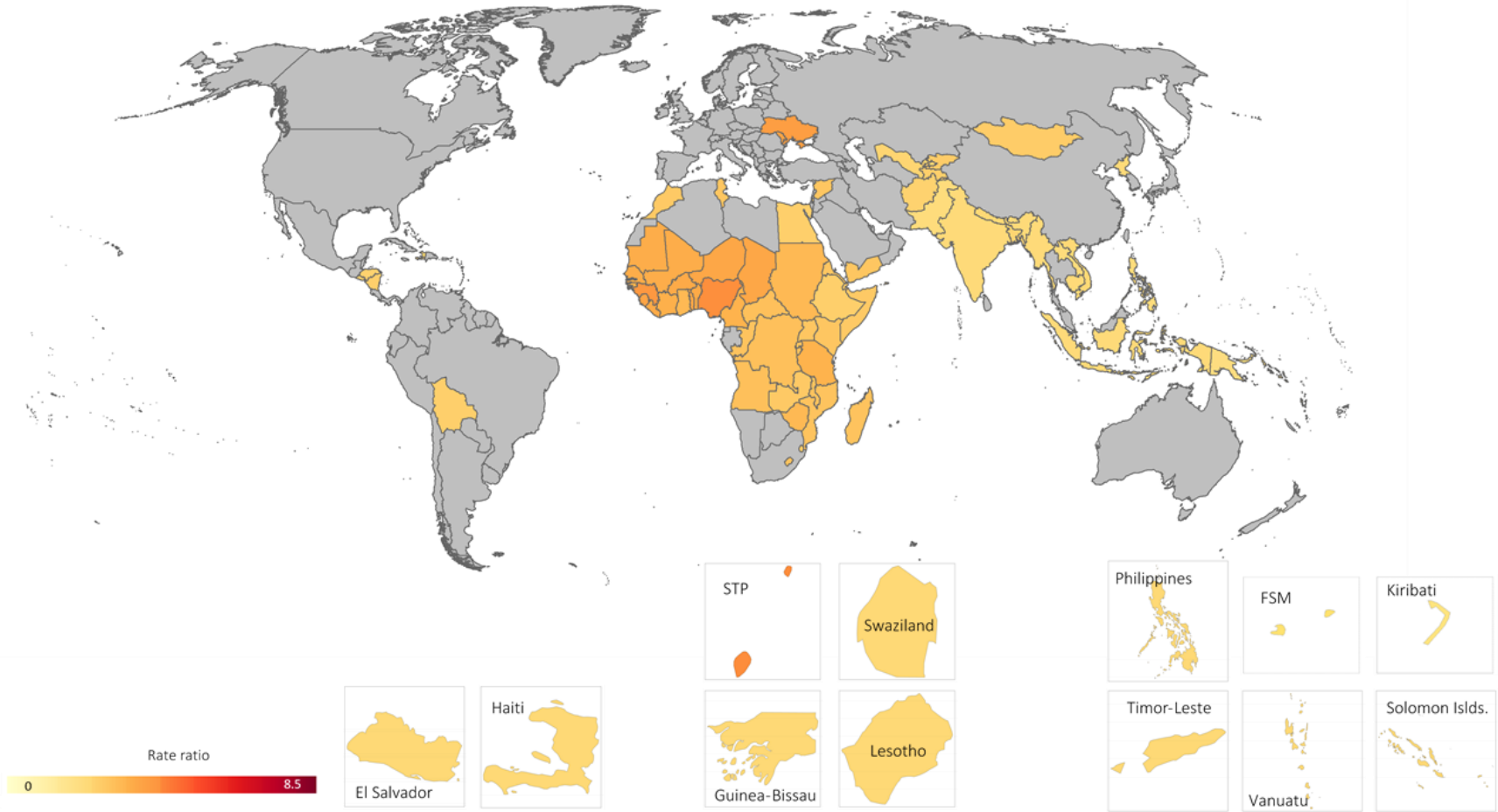


Figure 19d. Spatial distribution of the mortality rate ratio of non-diabetes-related CKD to diabetes-related-CKD for females 70+ years in lower income countries. Inserts highlight smaller countries; STP=Sao Tome and Principe, FSM= Federated States of Micronesia

3.4 Discussion

Overall, the GBD data analyses have highlighted areas which are known to have high burdens of both all cause CKD and CKD attributed to DM. The signals of "non-DM related CKD" – a proxy for CKDu in these exploratory analyses - were more varied and presented some interesting results. Excess mortality was detected in the existing and well documented hotspots in Central America and South Asia, but also in other regions which previously have had no reports, such as the Caribbean. However, these signals were not reflected in the rate ratios, suggesting that (in the case of the GBD data) these signals are not indicative of true CKDu as it would be expected that the higher rate ratio of Non-DM-related CKD: DM-related CKD would be mirrored in same countries as those which had the higher SMR values for "non DM related CKD". This could be due to the quality of CKD mortality data, and the relatively small proportion of the population with the "non-DM related CKD" outcome. These findings must therefore be regarded as preliminary and require further investigation.

For the analyses presented here, I grouped countries according to their income category to reflect the availability of RRT, and therefore likely quality of data available [154]. As CKD death rates are a conflation of CKD burden and RRT availability, there is significant variability in the capacity for renal care across upper- and lower-income countries, making a direct comparison between these groups challenging[154]. This approach does however have limitations, as countries within some regions – specifically across Central and Latin America - have very wide economic disparities, but only one central renal registry resulting in non-representative outcomes, especially for the lower income nations.

Although the patterns cannot be directly compared, there were similarities in the geographical distribution of excess mortality across high- and low -income countries. Excess mortality for both all cause CKD and CKD attributed to DM was generally inversely associated with age. For non-DM related CKD, excess mortality was significantly higher for males, and the highest mortality was observed amongst 15-49-year-old age group. This finding also concurs with the literature that CKDu disproportionately affects younger working age males[29, 61, 63, 163, 164].

Across both income groups there were regions which had consistently high excess mortality from all causes of death. In Oceania, the annual mortality rate due to CKD has increased at an average rate of 3.4% per year, of which 16% were due to DM-related CKD [165]. The age at which people develop end stage renal disease is considerably younger among indigenous peoples than among the mostly White nonindigenous people, where the mean age at onset of RRT is

44 years versus 60 years respectively [166]. This is believed to be due to earlier onset of coronary artery disease in these populations due to changing diet and lifestyle habits resulting from rapid economic transitions [165, 167]. There are an increasing number of CKD registries being established across the region, however for many countries, patients receiving dialysis and medicines represent the most privileged of the population. DM is the single greatest risk factor for kidney disease in Oceania and is estimated to be 2–5 times more common than in the primarily White, nonindigenous populations of New Zealand and Australia[168]. This is thought to be attributed to higher muscle mass and a higher prevalence of smoking amongst indigenous peoples[165]. Across the lower income nations in Oceania, the excess mortality for "CKD attributed to DM" was generally higher than those in the upper income countries. This could be due to factors including poor diet, smoking and obesity but could also be coupled with a higher quality of renal health care provision, and more sophisticated mortality data collection in comparison to the lower income countries in the region.

In both income groups there were strong signals observed for "non-DM related CKD". To date however, there are no reports of CKDu in Oceania, and it is likely these signals are reflections of the wider all cause CKD and DM-related CKD burden across the region.

There were interesting regional variations of excess mortality in each income group; specifically, in the detection of "non-DM related CKD". In the upper income countries, signals were detected in South East Asia and the Caribbean. In lower income nations there were signals in South Asia (specifically for females) and across Sub Saharan Africa. These variations will be discussed below by income category.

In high income nations, excess mortality for both all cause CKD and CKD attributed to DM was generally higher amongst females, particularly in the 15-49 age category. This finding concurs with a growing body of research into sex-specific mortality from CKD[169, 170], which shows notable differences which are attributable to socioeconomic and biological factors. The simplest explanation could be that females have a longer life expectancy than males, which, combined with the natural decline of kidney function with ageing contributes to an enlarged population dying form CKD[171, 172]. However, it is also possible that the equations used to estimate GFR result in CKD overdiagnosis in women because these equations, such as the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI), were not developed in community-based populations with appropriate age, sex and ethnic diversity[51, 173, 174], which

is discussed in chapters 1 and 5. Geographically, the highest excess mortality for all cause CKD and CKD attributed to DM was seen across Oceania (the Marshall Islands) and Southeast Asia (Mauritius), followed by Central America (Mexico) and the Caribbean (Dominica).
For non-DM related CKD, signals were markedly higher amongst males than females across all age groups, and excess mortality was inversely associated with age. Excess mortality was seen across the same countries as above, with the highest SMR observed in Mauritius followed by Mexico. Interestingly there was also a strong signal across the Caribbean, particularly amongst 15-49-year-old males, where the excess mortality from CKD attributed to DM was well below the global reference, suggesting this could be a potential area to look at in future.
The Caribbean is one of the world's most inequitable regions, with vast disparities between income and health profiles. The true burden of CKD is thought to be substantial, however, it is far from being established [175]. Since the development of the Caribbean Renal Registry (CRR) in 2006, there has been an ongoing attempt to advise Caribbean countries on the usefulness of the registry and the impact it can have on CKD treatment [176]. Recent reports suggest that hypertension and DM are the commonest causes of CKD amongst older people, and that dialysis is only offered in few of the islands. Of those who do receive RRT, the large proportion are males. Reporting to the CRR however is not mandatory, which likely results in an underestimation of the total number of individuals with on RRT and therefore CKD estimates[175]. To date there have been no reports of CKDu across the Caribbean Islands, which is likely the result of the lack of data in the region.
Neighbouring the Caribbean is Central and Latin America, where there is a quantified and significant burden of CKD[177]. Across the higher income nations, CKD [178] has similar etiological factors to those in the worlds most developed regions. To date, most reports of CKDu have come from low-income nations, predominantly in Nicaragua and El Salvador, where estimated prevalence rates range from 18% to 41% respectively [14, 71, 179]. Until recently it was unknown if CKDu extended north of El Salvador and Nicaragua, however, recent studies have suggested that there is an increasing prevalence of the disease in both Mexico and Guatemala[9]. These findings generally agree with the literature that CKDu is likely more widespread across Central American countries than previously expected, and thus warrant active detection studies in future. Retrospective dialysis enrolment data in Guatemala showed that the Southwest of the country, which is geographically contiguous

with El Salvador and Nicaragua, had the highest enrolment rates[9] a majority of which were males. Based on the contextual similarities with the broader region and on the high prevalence of young men among those undergoing dialysis treatment, it is hypothesized that CKDu is present in Guatemala[9, 180, 181], however, more research is needed to identify cases. Globally, CKDu is suspected to disproportionately affect working age males in agricultural communities[5, 13, 14, 29, 70, 164]. The disease does affect a smaller proportion of females who do not work in agriculture, however data on this population remain scarce as the focus is predominantly on working-age males[169, 182]. Amongst females, excess mortality for CKDu was detected in Mexico and Guatemala, which, for working age females was markedly higher than the excess deaths observed for "all-cause CKD". In general, however, the excess deaths for this group were markedly lower than those observed amongst their male counterparts.

The Latin American Society of Nephrology and Arterial Hypertension (SLANH) collects data on 20 national affiliates in the region and was initiated in 1991 and is widely used for prevalence and mortality estimates for CKD. Like the Caribbean, these countries have extreme socioeconomic and healthcare provision landscapes, making data collection a challenge. Thus far, the registry has been a valuable source of information regarding the epidemiology of CKD, however it has several limitations. First, most national registries in the region are voluntary and not all affiliates report annually, and some report only consolidated data. Second, many higher-income nations such as Mexico and Guatemala, report national estimates based on regional registries which limits the reliability of the estimates created using these data[183].

As a result of the economic and epidemiologic transitions experienced by many developing nations, risk factors for CKD have evolved [184, 185]. Although viral infections such as HIV account for a large proportion of CKD in developing nations, CKD is now increasingly a result of non-communicable diseases such as DM which was previously only seen in developed nations. These factors, coupled with poor healthcare, have rendered CKD increasingly more common in developing nations[60, 158].

For males in low-income countries, the highest excess deaths for all cause CKD and CKD attributed to DM were observed across Central America, Sub Saharan Africa and Oceania and India. For non-DM related CKD, the highest excess deaths were seen in among the 15-49-year olds in Central America, specifically in El Salvador which had a markedly higher SMR that the other regions. Significant signals were also seen in eSwatini in Sub Saharan Africa.

For females, the geographical distribution of excess mortality was slightly different. The highest SMRs for all cause CKD and CKD attributed to DM were seen in Oceania and South Asia – notably in Pakistan - and the lower excess deaths were in Central American countries. The SMR values for CKD attributed to DM increased with age, and for the 50-69 and 70+ ages, excess mortality was proportionally higher than the non-DM related CKD, especially suggesting that DM is the leading cause of CKD-related excess deaths in the older age groups.

Lower income countries across Central and Latin America comprise a significant CKDu "hotspot" [11, 28, 29, 182, 186]. As discussed previously, registry data for the region is fragmented. Differences in access to kidney care have been directly related to income, and a linear correlation between the gross domestic product and the prevalence of kidney disease has been described. Across the region, El Salvador and Nicaragua have some of the lowest expenditures on health per capita[187], and some of the highest prevalence and excess mortality from CKDu. In El Salvador and Nicaragua epidemiological knowledge of CKD is sparse, due to a lack of centralized data collection; however, from cross-sectional studies it is known that excess mortality from CKD is considerably higher in El Salvador and Nicaragua compared with other Central American countries. Furthermore, the male-specific CKD mortality rates far exceed rates among women in both Nicaragua and El Salvador[15, 188].

Signals of non-DM related CKD were also observed in Southern-sub-Saharan African countries, most notably in eSwatini. Currently, there are anecdotal reports of CKDu in North and West Africa[189, 190], however an inherent lack of high-quality mortality data limit definitive inferences of the disease burden at the national and regional level. A systematic review estimating prevalence of CKD in Sub-Saharan-Africa concluded that it is a prevalent and potentially escalating disease, with both communicable and non-communicable risk factors[191]. Infectious diseases - predominantly HIV - represent a unique cluster of risk factors responsible for CKD[192]. HIV is pandemic in Africa, with the highest prevalence rates estimated in Southern-Sub-Saharan Africa[193]. Lesotho and eSwatini have some of the highest HIV prevalence rates in the world (23.6% and 27.4% respectively)[194, 195], both of which have relatively high uptake rates of antiretroviral therapy[196, 197]. The success in bringing effective antiretroviral treatment to HIV-infected patients in Africa has determined the emergence of chronic medical illnesses such as HIV-related CKD[198]. It is therefore likely that the signal detected here is due to HIV-related CKD.

CKD in South Asia has the same pattern of widely varying prevalence as observed in developed countries. The causes of CKD are different and some of the regions have shown a high prevalence of CKD caused by glomerulonephritis and obstructive nephropathy and those of unknown aetiology. The majority of CKDu reports in South Asia are from Sri Lanka and India, which have been reported since 2002. Signals were observed across these countries, but these were comparatively lower than those in the other regions, which is an unexpected result, as will be discussed in chapter 4.

The strongest signal in South Asia for females was observed in Pakistan. The epidemiology of CKD and its risk factors are not widely studied here, but existing (sparse) data from cross sectional studies suggest approximately 25.3% of adults have reduced eGFR, of which only 2.3% were aware of their status. In this subset, 20% had coexistent diabetes[199], which is a factor which must be noted. Across South/East Asia it is estimated that more than 72 million adults have diabetes; which indicates that more than 60% of the world's population of people with DM will be in Asia[200, 201].

Amongst the females in the 15-49-year age group, CKD attributed to DM had very low excess mortality in comparison to all cause CKD however for the older females, the excess deaths increased significantly. This suggests that diabetic CKD is an issue for older populations in Pakistan, as concurred in the literature[202, 203]. In the case of CKDu in Pakistan, there have been some anecdotal reports from rural areas, however, no sex-specific data are available[204]. As CKDu is a growing public health issue in neighbouring India, Pakistan could be a potential at-risk country, as it relies on similar agricultural income, and in parts, has a similar climate to endemic regions.

These data are not granular enough to determine however whether the CKDu signal is indicative of true CKDu, or just a reflection of the diabetes epidemic in the country. It can be suggested that the passive detection method for CKDu is not appropriate due to the varying levels of data quality according to national income, RRT capacity and differing clinical practices when recording cause of death. The likely impact of applying this method to routinely collected data sources in suspected at-risk nations would be a broad misclassification of “non-DM related CKD” and spurious high SMRs in countries which in fact are not CKDu endemic. In turn, this could mis-direct cohort studies to countries which have spuriously high SMR values for the disease, whilst overlooking regions which could be at

significant risk for CKDu but have been overlooked using this preliminary method of detection.

3.4.1 Rate ratios

The ratio maps for both the high- and low-income countries provided unexpected results. The highest ratios of non-DM related CKD to DM related CKD were consistently seen in countries which had the lowest excess mortality for each cause of death, and which had no signals of non-DM related CKD in the main analyses.

For both males and females in the upper middle countries, the highest ratios were observed in Russia and Albania in addition to Saudi Arabia for females in the 15-49 category.

In these regions, RRT is largely available and free at the point of delivery (and thus a higher quality of mortality data are available), these high ratios do not reflect the signals seen in the SMR data which is likely due to the issue that the number of DM CKD and non-DM numbers are small for each region, resulting in spurious results which highlight stochastic fluctuations.

For both males and females in the lower-middle income nations, the highest ratios were observed in Africa, largely in Nigeria and Rwanda. The high ratios observed here however, are likely reflecting the high rate of HIV which is known to cause renal damage.

These rate ratios more likely highlight the fact that the CKDu signals detected in these countries are more likely a reflection of the high all cause and DM-related CKD in these regions as opposed to an actual detection of CKDu, and that passive detection using routinely collected data is not the best approach.

These aspects raise the question about how to effectively model these "small" data and account for the uncertainty in resulting estimates in future analyses. There is much discussion in the literature on the most appropriate methods to address the small numbers issue and surrounding uncertainty[205-207]. The SMR provides reliable estimates if there are large numbers of cases. However, this method can provide misleading estimates when the number of cases is small due to stochastic fluctuations. It has been suggested that empirical Bayes-estimated SMRs [205, 208, 209] may provide more reliable estimates of disease, particularly in areas which have a very low (or non-existent) prevalence of CKDu.

A prudent method would be to conduct classical SMR estimations and Bayesian estimations and compare the mean absolute error between the prediction models. Comparison maps could then be created to highlight the differences between the different models and their

respective errors in order to effectively communicate the uncertainty which different models can create.

3.4.1.1 Strengths and limitations

There are many strengths to the GBD dataset, namely that it is the largest and most sophisticated collection of mortality data in the world. This dataset has provided comprehensive estimates of the burden of disease in regions of the world which have sparse data, thanks to advanced statistical methods, and was the most appropriate data to use for this exploratory geographical analysis on the global scale.

The GBD data has important limitations. Methods for generating mortality metrics globally and by time period, country and region are complex and difficult to fully appraise. Moreover, the sources of data used in the GBD analyses require considerable assumptions, especially for countries where renal registry data are sparse. Furthermore, estimating a signal of CKDu from the all-cause condition depends on the validity of ICD coding, which requires accurate understanding of the coding system, which is likely to be lacking in both upper- and lower-income nations. Many of the regions which had high excess mortality currently have CKD surveillance systems which are under development or are non-existent which makes it difficult to estimate the true burden of this condition, and to determine the proportion of persons with CKD who could be classified as having non DM related CKD using routine data.

3.2.3 Conclusions

Routinely collected data are a valuable tool in the characterization and geographical distribution of the overall CKD burden. In these analyses, the distribution and burden of all cause CKD and CKD attributed to DM generally concurred with the existing literature, however, this was less accurate for non-DM related CKD. This is due to many factors but is most likely a result of the lack of renal registries across many regions, and coding disparities between clinicians when characterising cause of death resulting in misclassification bias. These issues make it challenging to draw conclusions about the distribution and causes of CKD mortality in a population. The passive detection method is therefore not a reliable way

to detect signals of non-DM related CKD at the regional or national level; and the active detection method with individual level data is more appropriate.

4. Geographical analyses of environmental risk factors for CKDu across Sri Lanka and India

4.1 Introduction

Results in chapter 3 showed that routinely collected data are a useful tool to some extent, for highlighting broad disease prevalence patterns across varying strata, especially in resource constricted regions. However, in the case of "non diabetes related CKD" (CKDu), this method is not appropriate due to potential deficiencies in data quality, and systematic errors resulting from complex coding systems[210] which could lead to erroneous prevalence estimates. As discussed in chapter 3, there are increasing reports of CKDu across South Asia, specifically in Sri Lanka and India; however signals of "non diabetes related CKD" from passive detection analyses in these countries were comparatively lower than other countries in the region which was an unexpected outcome, as CKDu has been reported in both countries in many studies across the literature[5, 64, 67, 211-213].

As a result of increasing CKDu reports in Sri Lanka and India, and the need to adopt standardized, internationally comparable data collection methods to measure kidney function, both countries have adopted the DEGREE protocol to study endemic regions (as introduced in chapter 2 and described in chapter 5). Sri Lanka recently completed a standardized cross-sectional study published in 2019, which estimated that the prevalence of CKDu in the population sampled was 6% which warranted further analyses assessing environmental risk factors [6]. India is currently conducting a DEGREE study after a descriptive paper which used data from national non communicable disease cohort studies, estimated that the prevalence of CKDu was 1.6% [5]. The findings from these studies are a key first step in measuring kidney function and estimating the prevalence of CKDu in South Asia. However, neither of these studies have included analyses of environmental risk factors in association with CKDu, which will facilitate international comparisons of these risk factors.

In this chapter I analyse environmental risk factors for suspected CKDu (characterized by eGFR <60ml/min/1.73m^2) in Sri Lanka and India using cross-sectional data from recently published studies.

As detailed in chapter 2, there are several environmental risk factors which are hypothesised as related to CKDu which are assumed to be common across endemic regions because the disease is largely reported in hot, low-lying agriculture-dominated regions. Due to a lack of

environmental data collection and mapping in relation to these risk factors however, it is not clear whether these factors account for the population patterns of CKDu.

I conduct descriptive exploratory analyses to assess associations between eGFR as a continuous variable and risk of eGFR<60 with the leading hypothesised environmental risk factors: (i) ambient temperature [29, 31]; (ii) altitude [14, 214]; (iii) proximity to agricultural land [10], which is increasingly being studied as a proxy to agrochemical exposure due to resource and financial constraints of effectively measuring pesticides in laboratories

To investigate possible associations between these environmental risk factors and eGFR I conducted the following:

(i) Use cross sectional study data from reported at-risk regions of Sri Lanka and India

(ii) Assign satellite derived environmental risk factor variables heat index, altitude, and proximity to land cover (as a proxy for exposure to pesticides), to participant housing coordinate data across both study countries

(iii) Investigate potential associations between the purported environmental risk factors and eGFR

As discussed above, Sri Lanka is a DEGREE study centre which surveyed kidney function following the standardized core protocol. Participants were sampled from the North Central Province (NCP) which is considered a CKDu hotspot [25, 30, 215]. For the Indian dataset, the Public Health Foundation of India provided data from two government funded cohort studies conducted in northern and southern India across a mix of urban and rural areas. These data were first used by researchers at the London School of Hygiene and Tropical Medicine who harmonized these data in order to conduct descriptive analyses on the presence of CKDu in these study zones [5]. As these studies have used different measurement methods for kidney function, prevalence data are not strictly comparable. However, as there is no clinical consensus as to whether CKDu is the same disease across these regions, and because my analyses investigate broad spatial and environmental patterns in relation to eGFR, the datasets will be compared and discussed in relation to environmental risk factors only.

Therefore, the objectives of this geographical analysis are to assign environmental variables to study participants in the Sri Lankan and Indian populations using satellite and reanalysis (retrospective modelling) datasets. Also to assess potential associations between low eGFR and risk of eGFR<60, and the key hypothesised environmental risk factors of ambient temperature [29, 31] altitude [14, 214] and proximity to agricultural land cover [10]. Finally, I assess whether these risk factors are uniform across these endemic regions, as has up until now been assumed in much of the CKDu literature.

4.2 Methods

4.2.1 Study settings, participants, study designs and health data

4.2.1.1 Sri Lanka

Sri Lanka is an island nation in the Indian Ocean located on latitude 6.00°N and longitude 81.00°E. The country has a predominantly agricultural economy with a majority of the population living in rural areas (77.4%, correct as of the 2011 census) [216]. Urbanization in Sri Lanka is lower than many other Asian countries, with an estimated annual growth rate of 1.5% in 2018 [217]. The DEGREE surveys in Sri Lanka were conducted in five geographically demarcated agricultural communities in the Anuradhapura District in the North Central Province (NCP) between March and May 2017 (figure 20).

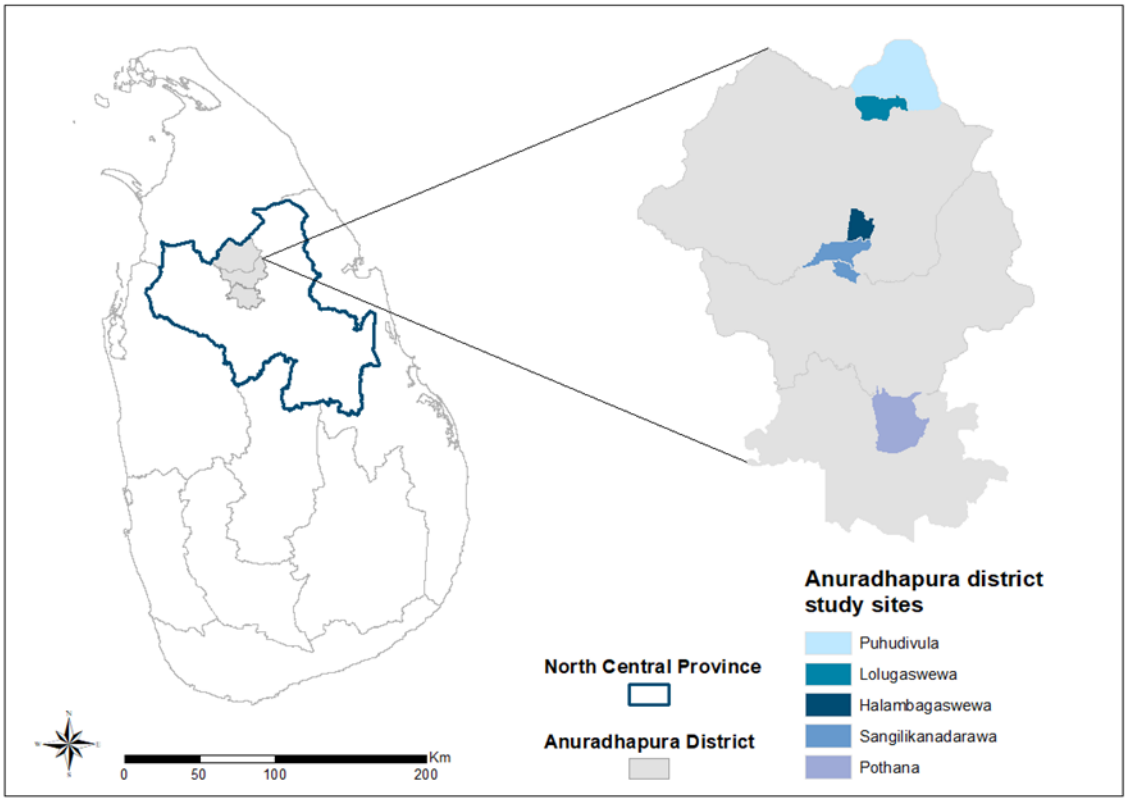


Figure 20. Map of Sri Lanka showing the five communities selected for the DEGREE survey in the North Central Province, Sri Lanka.

[Shapefiles downloaded from the DIVA GIS [218] shapefile database at administrative boundary area level 2 (see appendix 1, figure 1 for administrative area levels). The shapefile is projected into the world geodetic WGS1984 coordinate system]

These five communities were selected for the study as there was information from government screening programmes which have previously classified these districts as “low”, “moderate” and “high” for CKDu prevalence. In figure 21 I have mapped the Government classified study zones which were used as the analysis frame for the DEGREE study.

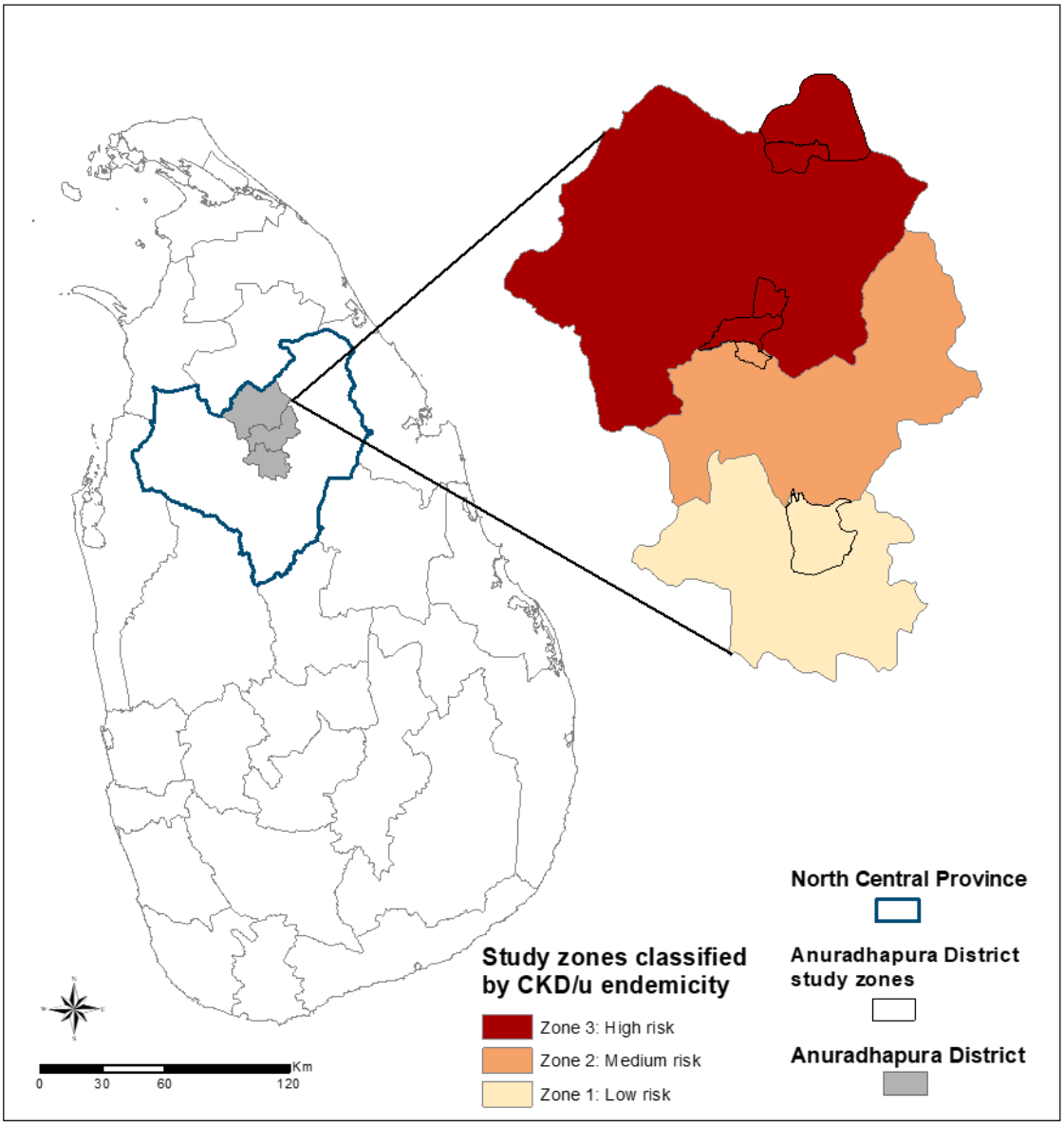


Figure 21. Map of Sri Lanka DEGREE study zones grouped according to governmental classification of CKD/u endemicity.

[Shapefiles downloaded from the DIVA GIS[218] shapefile database at administrative boundary area level 2. The shapefile is projected into the world geodetic WGS1984 coordinate system]

All adults over 18 years of age with a permanent residence in the study area were invited to participate in the study. Using a voters list, all households in the study area were visited and eligible adults were invited to enrol. The housing coordinates of all consenting participants were recorded using a global positioning satellite (GPS) device. Consenting participants were interviewed at their residence using the DEGREE core questionnaire and were also given a short environmental questionnaire to collect data on variables including pesticide use and

drinking water sources. The day after interview, participants attended a local clinic to provide field clinicians with fasting blood and urine samples. Anthropometric measures were also taken during the same clinic visit. Full details of data collecting procedures and methods are described elsewhere [6].

I used data from a community-based cross-sectional household survey which was conducted amongst 11,217 adult participants over 18 years of age in five communities in the North Central Province in 2017 using the DEGREE protocol. Participants were interviewed using a core questionnaire to obtain data on age, sex, income, schooling years, alcohol intake, smoking, meat consumption and medical history and occupational status. Anthropometric measurements including body mass index and fat-free mass were taken, as well as blood and urine samples. For full protocol description and study objectives, see chapter 5.

4.2.1.2 India

The Indian subcontinent is located on latitude 21°N and latitude 78°E. Similar to other CKDu-endemic countries, India has a largely agricultural-led economy, and a majority of the population are rural (73%) [219]. The Indian cohort studies used in this chapter are described below.

Implementing a Comprehensive Diabetes Prevention Management Program (UDAY) study

The UDAY study is a 5-year initiative, which aims to reduce the risk of diabetes and hypertension and improve management by implementing a comprehensive intervention programme in urban and rural sites. This study has a pre-post evaluation design with representative cross-sectional surveys before and after intervention. A total of 100,000 adults over 18 years of were selected for baseline and postintervention assessments [220]. I used data from the first cross-sectional study which was conducted between July-December 2014 in urban and rural sites in the States of Haryana in Northern India, and Andhra Pradesh in the South. In Haryana state, both the urban and rural sites were located in the district of Sonipat which lies north of the Capital city of New Delhi. The urban site was in Sonipat city, which covers a 388 Km^2 area, and has a population of approximately 1,450,000 [219]. The rural site was in the Kharkhoda sub-district which covers 278 Km^2 with a population of 1,60,895 (correct as of 2011)[219] (see figure 21). The UDAY sites in southern India were located in the state of Andhra Pradesh state on the South East coast in the Visakhapatnam district (Vizag). Covering

11,161 Km^2, with a population of 4,290,589 (correct as of 2011)[219], Vizag has 43 sub-districts comprising 4 urban and 39 rural areas. The urban site was located in the city of Visakhapatnam, and the rural site comprised the sub-districts Makavarapalem and Nathavaram [220] (see figure 22).

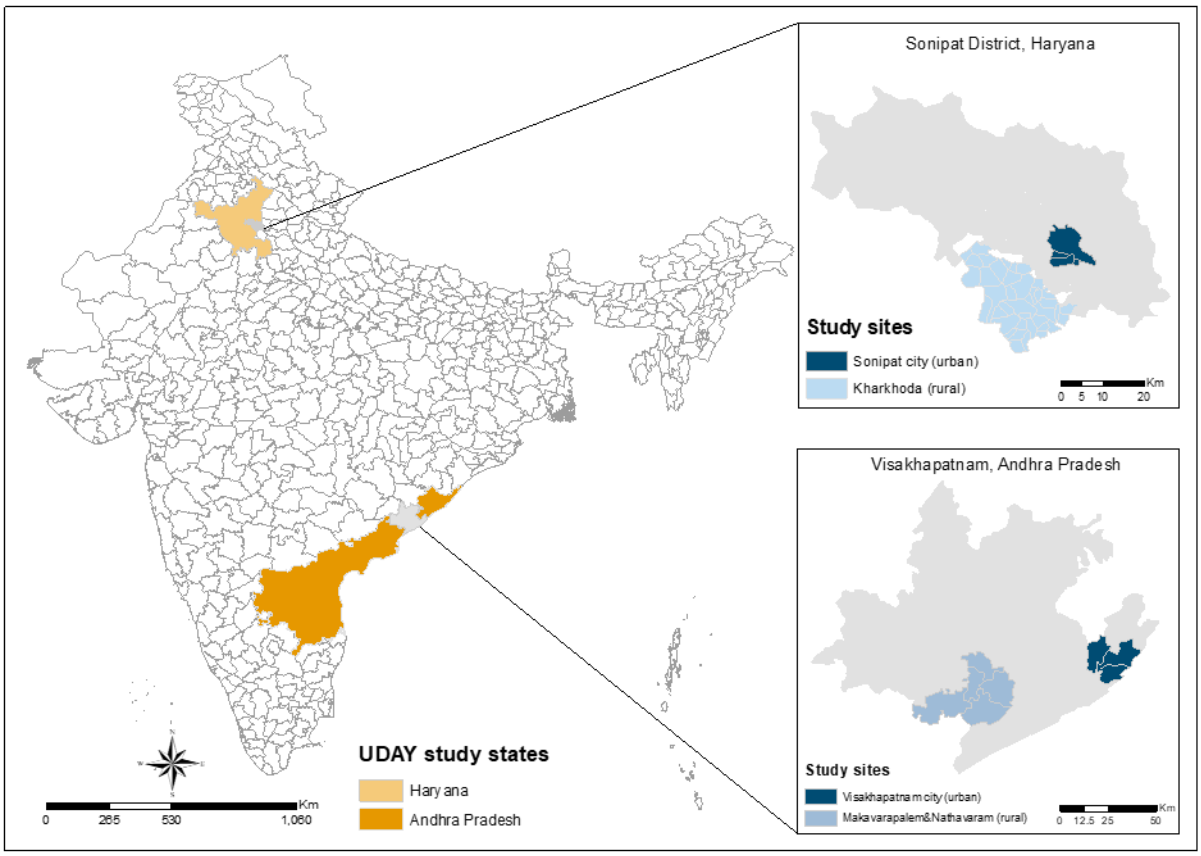


Figure 22. Map of India showing the "*Implementing a Comprehensive Diabetes Prevention Management Program"* (UDAY) study sites located in rural and urban Sonipat (Haryana) and Vishakhapatnam (Andhra Pradesh).

[Shapefiles downloaded from the ArcGIS shapefile database [221] at the sub-district level (Tehsil) (see appendix 1, figure 2 for administrative area levels), and was projected into the world geodetic WGS1984 coordinate system]

Centre for Cardiometabolic Risk Reduction in South Asia (CARRS) study

The CARRS study was conducted between 2010-2011 in two urban sites in North and South India. The northern site was located in India's capital city New Delhi which covers a 42.7 Km^2 area and has a population of 142,004 (correct as of 2011)[219]. The southern site was in Chennai, the capital of Tamil Nadu state (Figure 23). Chennai covers an area of 426 Km^2 and

has a population of 7,800,000 [3]. The CARRS study is a hybrid model with two cross-sectional serial surveys three years apart to monitor the incidence of cardio metabolic diseases, with a three-year prospective follow-up of the first cohort. The study was conducted in three urban settings (Chennai and New Delhi in India; Karachi in Pakistan), with 4000 participants in each site stratified by gender and age. In this chapter, I used data from Chennai and New Delhi only. Participant sampling was conducted using a multi-stage cluster random sample framework, followed by within-household participant selection through a combination of Health Information National Trends Survey (HINTS) and Kish methods[222].

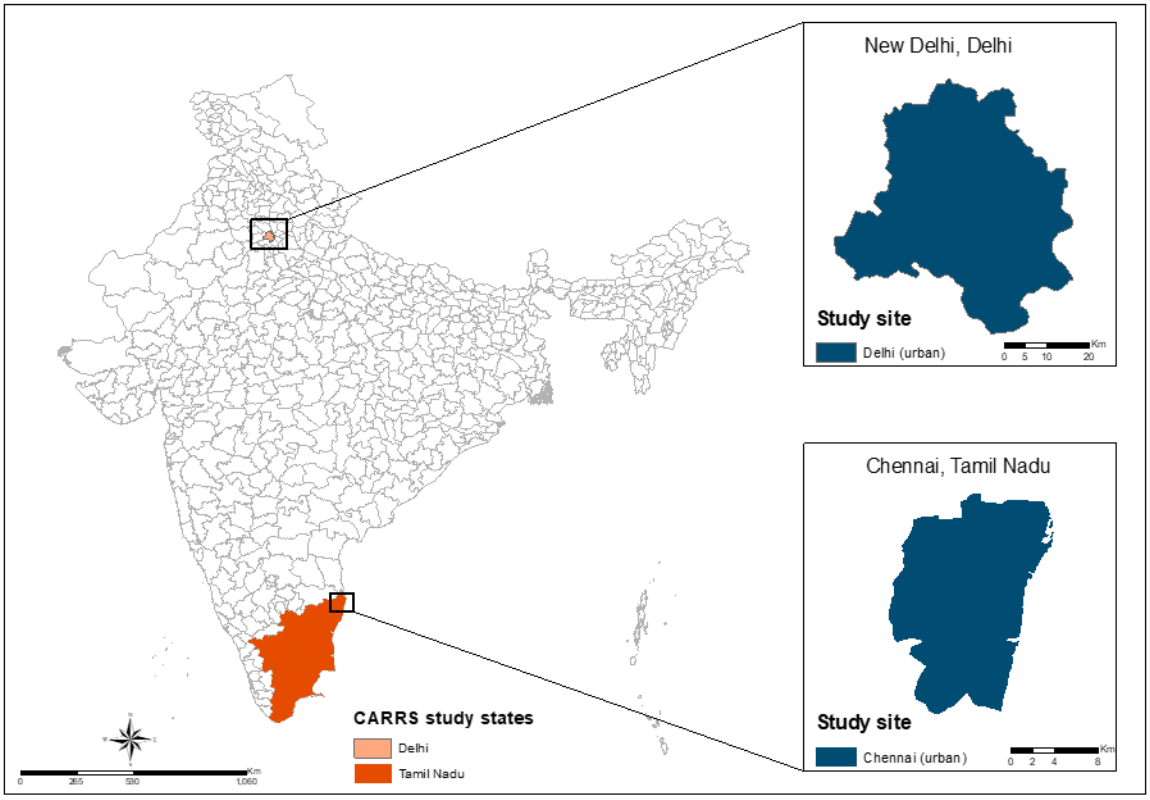


Figure 23. Map of India showing the "*Centre for Cardiometabolic Risk Reduction in South Asia"* (CARRS) study sites located in urban Delhi (Delhi) and rural Chennai (Tamil Nadu).

Shapefiles downloaded from the ArcGIS shapefile database[221] at the sub-district level (Tehsil), and is projected into the world geodetic WGS1984 coordinate system

Across both studies, the participant housing coordinates were recorded using GPS devices, at baseline. Both studies were led by the Public Health Foundation of India in collaboration with

a consortium of public health and population services research bodies [220, 222]. See table 16 for an overview of the study designs for the Indian datasets.

Table 16. Design and recruitment methods for the studies used in the India CKDu analysis

	CARRS		UDAY			
Study period	October 2010 – November 2011		July 2014 – December 2014			
Geographic region	North	South	North		South	
Urbanicity	Urban		Urban	Rural	Urban	Rural
State	Delhi (state of Delhi)	Chennai (state of Tamil Nadu)	Sonipat (state of Haryana)		Vishakhapatnam (state of Andhra Pradesh)	
Household sampling	Multistage cluster random (ward-census-enumeration block-household)		Multistage cluster random (Census Enumeration blocks (urban) or villages (rural) – households)			
Individual sampling	1 man 1 woman from each household (Kish method used, (World Health Organisation, 2015)		1 man 1 woman from each household (Kish method used, (World Health Organisation, 2015)			
Consent prior to participation	Yes					
Age groups included in analysis	≥20		≥30			
Exclusion criteria	Pregnant, bedridden and participants who were unable to comprehend the questionnaires due to cognitive deficiencies					
Laboratory[a]	PHFI[b]	MDRF[c]	PHFI[b]			

[a]Study laboratories participated in Randox International Quality Assurance Scheme (RIQAS) for clinical chemistry an HbA1c during the study periods; [b]Public Health Foundation of India; [c]Madras Diabetes Research Foundation

4.2.1.3 Environmental Data

The growth in the use of remote sensing and geographic information systems (GIS) in public health has facilitated the analyses of multiple environment-disease associations at varying geographical resolutions [223-226]. Satellite measurements have distinct advantages over conventional ground measurements because data can be collected repeatedly and automatically and can provide better coverage than ground monitors.

In order to assign participants with environmental variables to facilitate regression analyses, I modelled satellite-derived heat index temperature, land cover and altitude data to create continuous environmental surfaces over Sri Lanka and India. I assigned a value for each variable to each participant residential coordinate to facilitate regression analyses between eGFR and these environmental variables, each of which is a postulated risk factor for CKDu. Table 17 shows a summary of the satellite data sources I used in these analyses.

Sri Lanka included additional data on self-reported environment-linked lifestyle factors drinking water sources, use of pesticides, and types of agricultural crops cultivated as a proxy for potential pesticide use; all of which could influence eGFR.

Table 17. Summary of satellite-derived environmental data sources used in this analysis

Risk factor	Data source	Spatial resolution	Time	Coverage
Meteorology (temperature, humidity)	ECMWF Era-5 dataset [227]	31 x 31 km 2m vertical accuracy	1979-present	Global
Altitude	Shuttle Radar Topography Mission [228]	30 x 30m 20m vertical accuracy	2000	Global
Land cover including crops and land use	European Space Agency Climate Change Initiative Land Cover (ESA-CCI LC) [229]	300 x 300 m	1990-2015	Global

In the following section I describe the environmental variables I used in this chapter, and detail how the data was processed and assigned to the participants residential coordinates in Sri Lanka an India.

4.2.1.4 Meteorological data

Across the CKDu research field, heat stress as a result of physical exertion in high ambient temperatures and humidity is a central aetiological hypothesis [13, 29, 31]. In order to estimate the combined effects of temperature and humidity on the human body, I used the heat index

(HI) as my primary meteorological outcome variable [230-233]. The HI is a commonly used exposure proxy for heat stress in environmental health studies [234-236] as it provides a "feels like" heat measure and is considered more robust than measuring air temperature alone to estimate heat exposure [232, 233].

There are many meteorological data sources which allow in-situ measurement of temperature, but station density is much higher in developed nations in comparison to low and middle income nations [237]. Other measurement platforms including aircraft and balloons are a good source for short-term weather predictions, but are limited due to changes in systems of observation and assimilation methods [227]. The creation of long-term global climate datasets has been a central international aim in understanding long-term large-scale climate patterns. Since the 1990s retrospective analysis (or reanalysis) has provided a successful method of assimilating data points from multiple sources into a single model output. I used reanalysis data developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), a research institute providing global numerical weather predictions. The ERA-5 reanalysis dataset is the latest global climate reanalysis produced by the ECMWF (2019), which has a 31km spatial resolution and hourly time resolution dating back to 1979. I used retrospective ERA-5 reanalysis to extract 2-metre height temperature and 2-metre height dew point variables for March-May 2017 (Sri Lanka DEGREE survey period) and October 2010 – November 2011 and July 2014 – December 2014 for the Indian study periods.

The HI is calculated using the following equation:

Equation 3. Heat index formula

$$\mathrm{HI} = -42.379 + 2.04901523\mathrm{T} + 10.14333127\mathrm{R} - 0.22475541\mathrm{TR} - 6.83783 \times 10-3\mathrm{T2} -5.481717 \times 10-2\mathrm{R2} + 1.22874 \times 10-3\,\mathrm{T2R} + 8.5282 \times 10-4\,\mathrm{TR2} - 1.99 \times 10-6\mathrm{T2R2}$$

[238]

Where, HI = heat index; T = ambient temperature; R = relative humidity

As the relative humidity is not calculated as part of the ERA-5 dataset, I downloaded monthly average files for ambient temperature and dew point temperature from ERA-5 which were

used to calculate the relative humidity, both of which were measured 2 metres above the land surface in Kelvin units on a gridded global dataset. Each grid contained a monthly average of recorded ambient temperature and dew point temperature, which I converted to Celsius. As the data is packaged on a global scale, I used GitHub country bounding box coordinate data to create a subset of temperature and dew point data for Sri Lanka and India which formed my climate datasets for analysis. Using the R studio "Weathermetrics" package I calculated the relative humidity using the gridded temperature and dew point data, followed by the heat index for each grid point in Celsius units.

To model a continuous heat index surface across Sri Lanka and India, I used the probabilistic interpolation method Empirical Bayesian Kriging (EBK). Kriging is a class of statistical techniques for optimal spatial prediction which assumes spatial dependency (or spatial autocorrelation) and stationarity[239]. Spatial autocorrelation is a measure of the degree to which observations (points) at specific spatial locations are similar to each other across a study area. The assumption of stationarity is that the joint probability distribution does not vary across a study area, and parameters such as the overall mean of values do not vary across space.

For geostatistical modelling, the structure of spatial variation is estimated through the semivariogram which is a visual depiction of the covariance exhibited between each pair of points in the sampled data. The semivariogram determines the weights to apply to data points when developing predictions and is thus a key step in kriging[240]. The semivariogram takes all possible pairs of measurements in the dataset, and then the squared differences between values of paired locations are calculated. Depending on sample size, the point pairs are usually 'binned' into groups with a similar distance (lag) and direction between them. The average semi-variance in each bin is plotted on the y axis (as can be seen in figure 24 below). To aid visualisation of the model fit to the data points, bins are further averaged and represented as blue crosses on the semivariogram. Semi-variance is expected to be lower with smaller distances between pairs of points.

The variogram is expressed as:

Equation 4. Variogram model equation

$$\gamma(h) = \frac{1}{2} Var\left\{\left(Z(s) - \mu(s)\right) - \left(Z(s+h) - \mu(s+h)\right)\right\} \quad (240)$$

Where: $\boldsymbol{h}$ denotes the translation between any two sites $\boldsymbol{s_i}$ and $\boldsymbol{s_j}$ in a study area.

The semivariogram has three parameters: the nugget, range, and sill. The nugget refers to variations at shorter distances within the dataset, which can be attributed to measurement error and microscale variation[241]. The range is the distance in which the semivariogram reaches the sill or levels off. Data points beyond the range are spatially independent and there is no spatial autocorrelation, and the sill is the value on the y-axis where the range levels off [240].

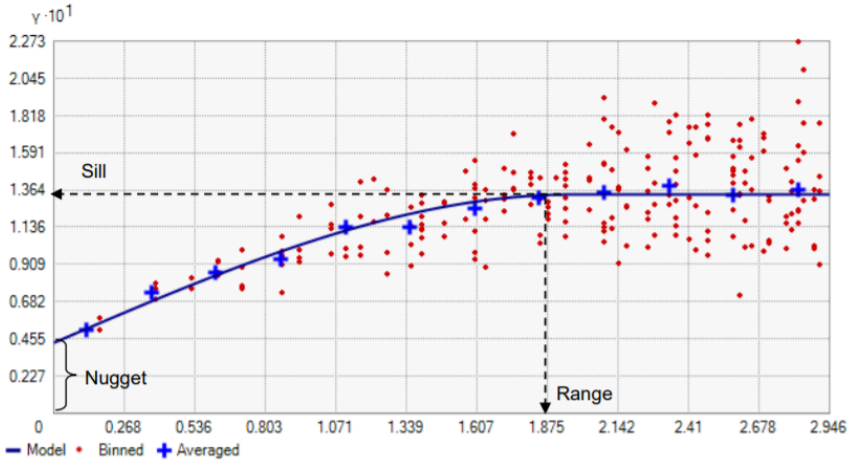


Figure 24: Diagram of a Semivariogram showing the main components (i) nugget (ii) range and (iii) sill. X axis shows the distance (lag) from the measuring point, Y axis shows average variogram value Source: adapted from ESRI [242]

Classical kriging assumes that the estimated semivariogram is the true semivariogram of the data, meaning data is generated from a Gaussian distribution with the correlation structure defined by the estimated semivariogram[239]. As environmental processes are highly variable across space, this method could result in underestimation of the standard error of the prediction due to overlooking the uncertainty of the semivariogram.

Empirical Bayesian kriging differs from classical kriging methods by accounting for the error introduced by the estimation of the semivariogram. This is done by estimating and re-using many semivariogram subset simulations, to create a spectrum of semivariograms rather than a single model [243, 244]. This method also does not assume a tendency to an overall mean, thus

allowing more flexibility in data deviations as seen in figure 25. Empirical Bayesian kriging therefore produces accurate predictions of moderately nonstationary data and standard errors of predictions are considered more accurate than other kriging methods [244].

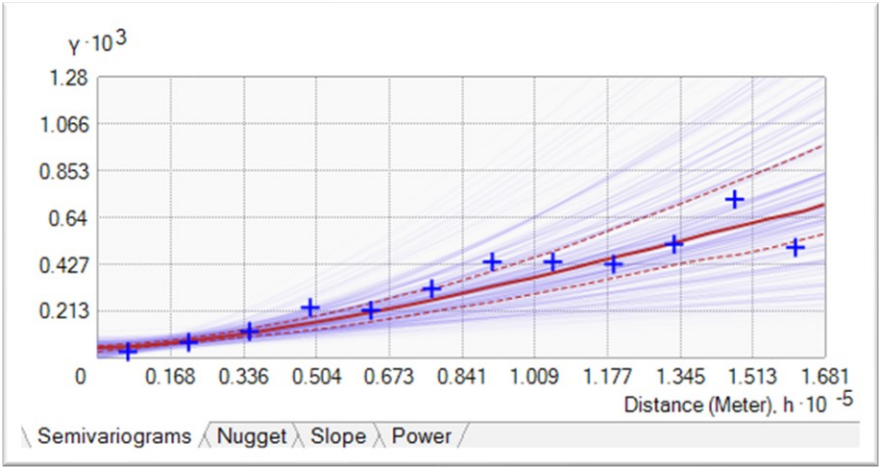


Figure 25: Spectrum of simulated semivariogram models produced using the Empirical Bayesian kriging method. X axis shows the distance (lag) from the measuring point, Y axis shows average variogram value

Before conducting the kriging analyses, I tested each heat index dataset (across Sri Lanka and the two study time periods in India) for normality of distribution, and stationarity. As each data set was non-stationary and non-gaussian (as is common in environmental variables), a more robust model was required to estimate the unknown covariance function, and therefore Empirical Bayesian kriging was selected as the optimal method to use. See appendices 2 and 3 for heat index distribution plots for Sri Lanka and India.

As the study zones across Sri Lanka and India were generally smaller than the 31km resolution of the ERA-5 dataset, I modelled the heat index surface at the country level, and then extracted the study zone shapefiles when assigning heat index values to participant coordinates.

Due to the skew observed in the heat index data, I log transformed the data in both Sri Lanka and India. Furthermore, in India there were negative heat index values, so I therefore added a constant to the values and log transformed the data as shown by:

Equation 5. Log transformation adding a constant to negative heat index values

$$Heat\ index = log(Y + a - min\ (Y)) \quad (245)$$

Whereby Y is the negative datum; *a* is the constant (1.001).

When determining the semivariogram and selected the default search neighbourhood to 1-sector and 10-15 nearest neighbour approach as comparatively this produced the best fit outcomes in comparison to other neighbour scenarios. Surrounding observations are weighted using the least squares method to make predictions for each location in the study area and selecting a search neighbourhood determines the distance at which points furthest away have less influence, and therefore less weight in the prediction models.

After the models were fitted, I conducted a validation of each model using a holdout cross validation approach by which the dataset is randomly divided into a training and validation set (or a hold-out set.) The model is trained on the training dataset and evaluated on the validation dataset[246]. The overall fit was evaluated on the basis of the route mean squared error (RMSE) and the mean actual error (MAE).

Finally, before using the heat index values in regression models, I took the antilog of the data.

Sensitivity analyses for heat index data

Spatial interpolation methods have been applied in many environmental research studies; however, it can be complex to select an appropriate method for the conversion of raw data (either from discrete measurement sites or reanalysis grids) into continuous maps.

In order to assess the gains from using the Empirical Bayesian Kriging method, I conducted a sensitivity analysis by comparing the original heat index data (and the proposed EBK model) with more simplistic interpolation methods: inverse distance weighting, bilinear interpolation and kriging. I selected a smaller search radius of 4 neighbours, and restricted data to altitudes within 100m of those in the target study zones across India and Sri Lanka. Model fit was compared using the hold out validation method, followed by a Pearson correlation analysis to help ascertain which interpolation method was likely the most appropriate.

Inverse distance weighting

Inverse distance weighting (IDW) is a deterministic interpolation method by which estimates are made based on nearby known locations. The weights assigned to the interpolating points (or remodelled data in the case of the heat index dataset) are the inverse of its distance from the interpolation point. Consequently, closer points have higher weights (and thus more impact) than distant points and vice versa. The known sample points are implicit to be self-governing from each other[247].

Equation 6. Inverse distance weighting equation

$$\hat{v} = \frac{\sum_{i=1}^{n} \frac{1}{d_i} v_i}{\sum_{i=1}^{n} \frac{1}{d_i}} \quad (247)$$

$\hat{v}$ = value to be estimated
v_i = known value
$d_i \ldots, d_n$ = distance from the *n* data points to the point estimated *n*

Kriging

The kriging method incorporates statistical properties of the measured data (spatial autocorrelation). The kriging approach uses the semivariogram to express spatial continuity (autocorrelation). Like EBK, kriging relies on the semivariogram which measures the strength of the statistical correlation as a function of distance (see figure 24 above for details on the semivariogram)[239].

Bilinear interpolation

Bilinear interpolation is a resampling method used for calculating values of a grid location based on the value of the four nearest input cell centroids to determine a value on an output raster. The output is a smoothed raster grid as a higher resolution than the original working environment[248].

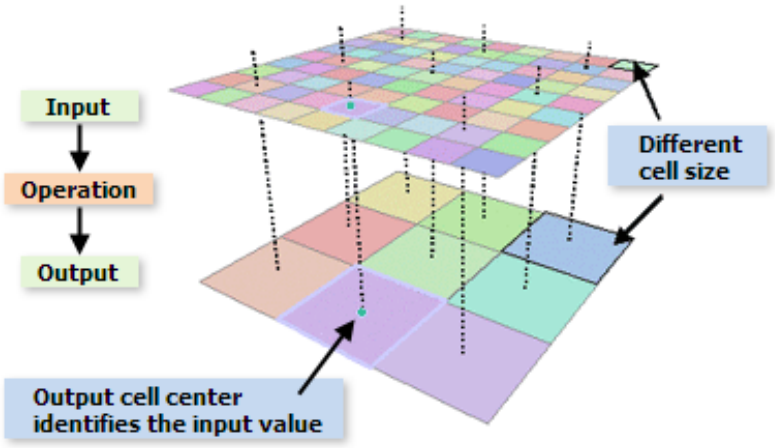


Figure 26. Infographic describing the resampling method for bilinear interpolation. (Source ESRI[248]).

In order to understand whether the choice of model affects the covariates in the regression models, I conducted modelling with the EBK heat index data and the bilinear interpolation heat index data with the aim of highlighting the importance of selecting the most appropriate model for predictive surfaces in environmental exposure analyses.

4.2.1.5 Digital Elevation Model data

Living and working at low altitudes is another hypothesised risk factor for CKDu, due to higher humidity levels, which can exacerbate heat stress, as discussed above. In order to assign an altitude value to each study participant, I used a digital elevation model (DEM) of the study zones. A digital elevation model is a 3D representation of bare ground derived from remote sensing data and is the most common basis for digitally-produced relief maps[249].

For India and Sri Lanka I downloaded satellite data from The Shuttle Radar Topography Mission (SRTM), a global DEM giving coverage of void-filled data at a resolution of 30m with a vertical accuracy of 20m [228]. The digital elevation data are available in raster image "tiles" which contain altitude data in 1-spectral band. I selected the tiles overlaying Sri Lanka and India and loaded them into GIS software ArcMap 10.5.1. Here the tiles were merged, or "mosaiced" to form a continuous elevation surface which were then clipped to each country shapefile.

4.2.1.6 Land cover data

Exposure to pesticides is a major topic of research and debate in the CKDu field, particularly in Central America[71, 250]. In India and Sri Lanka, pesticide exposure research is gaining traction,

however, it is still in its infancy. As there was little to no information collected on pesticide exposure in the Sri Lanka and India studies, I used residential proximity to cropland as a proxy for exposure to pesticides using satellite derived land cover class data.

Land cover data provides spatial information on different classes of physical coverage of the Earth's surface, including forests, croplands, and urban areas. Since the advent of remote sensing land cover classification in the early 1970s, data resolution and reliability have improved substantially, allowing for spatially explicit modelling in health research.

The European Space Agency (ESA) launched their Climate Change Initiative (CCI) programme aiming to provide high quality satellite-derived land cover products which are available on an annual basis from 1992-2015 at a 300m resolution. Unlike many remote sensing products that are based on single-year single-sensor approaches, this dataset is generated using multiple sensors with an overall accuracy of approximately 71% [251], with typology comprising 22 classes defined using the UN Land Cover Classification System. For global and regional legends, see appendix 4. I downloaded the global dataset from the ESA-CCI database in a raster format (a pixelated image), which contained values corresponding to a land cover category. I transferred the global dataset into ArcMap and overlaid a shapefile of Sri Lanka and India onto the raster image. I then extracted the land cover data within the shapefile of Sri Lanka and India.

4.2.1.7 Assigning environmental factors to participants

Using ArcMap, I overlaid the participant residence coordinates in the CKDu datasets for each country onto their respective shapefiles. The datasets contained coordinates of the participants' residence, which were used as the environmental data assignment points. Coordinates were transformed into The World Geodetic System 1984 (WGS84) coordinate system, which is the reference coordinate system used in global positioning systems (GPS) [252], and ensures that the coordinates and shapefiles are in the same datum.

Heat index and altitude values

Heat index and altitude raster data were assigned to each residential coordinate using the "raster to point value" extraction tool in ArcGIS.

As there were two sampling time points for the CARRS and UDAY studies conducted in India, I created two averaged heat index surfaces for each time period (October 2010 – November

2011 and July 2014 – December 2014 respectively) and then assigned values to CARRS an UDAY participant coordinates. I then merged the study datasets into a single dataset.

Land cover

To capture land cover in the immediate neighbourhood of participants and assign each a land cover class, I placed a 300m buffer (circular radius) around each residential coordinate to match the land cover data resolution. Using the spatial join tool in ArcGIS, I assigned the mode land cover class which fell inside each residential buffer. I selected the mode classification method as many residential coordinates were located at a cross-section of different land cover classes, and I wanted to capture the most common class for each participant.

4.2.1.8 Statistical Model building and analyses

In order to explore the associations between CKDu and its postulated risk factors, I modelled kidney function (eGFR) as a continuous variable with socioeconomic, anthropometric and environmental variables using linear regression models in both Sri Lanka and India. I used Logistic regression models to estimate odds ratios (ORs) and corresponding 95% confidence intervals (CI) for risk of having a low eGFR (<60ml/min/1.73m^2).

I tested each linear and logistic regression model for multicollinearity, as I suspected that some variables (especially the environmental variables heat index and altitude) would be correlated with each other. I used the variance inflation factor (VIF) >10 as the threshold for exclusion. Due to multicollinearity between heat index and altitude, I modelled these variables separately in the linear and logistic regression models.

Multiple regression models included all variables that had shown independent associations with eGFR in univariate analyses, coupled with all environmental variables as these were of a priori interest.

Having ascertained the key risk factors for low eGFR and eGFR<60 I then wanted to assess the effects of environmental variables on eGFR alone, accounting for different geographical locations. To do this I used Linear mixed models (LMM) to model the trends between eGFR and environmental variables between different geographical locations. Linear mixed models assume that observations in the same cluster (or study zone) are more correlated than those in another study zone [253], that the average eGFR will differ between zones, and that the effect of environmental variables on eGFR is stable across the study zones.

For Sri Lanka I modelled the available environmental factors heat index, altitude, land cover, pesticide use, chena cultivation, rice and vegetable farming and water source with study zones 1-3 as the random structure. I could not delineate this further into urban/rural areas as the entire area is a rural zone.

For India I modelled environmental variables heat index, altitude, and land cover. In each model the fixed effects were eGFR and the satellite derived environmental variables above. For the random structures I modelled state with the sub-category urban/rural area because the study zones across India are very large and require more delineation for meaningful results. I calculated a fixed effects model which illustrates how each environmental variable affects eGFR across the study zones. Finally, I compared environment-eGFR effects between Sri Lanka and India modelling the common environmental variables heat index, altitude, and land cover. The random structures were modelled in a hierarchy of country > study zone (Sri Lanka)/State (India). I used caterpillar plots instead of forest in order to illustrate the environment-eGFR interactions for each study zone as the items are sorted in order of increasing effect size and would be easier to interpret. For all models discussed above I modelled the variables using the heat index data derived from empirical Bayesian kriging methodology and the bilinear interpolation methodology to highlight the importance of selecting the most appropriate interpolation method to create predictive surfaces.

All statistical analyses were conducted in R Studio version 3.5.1.

4.2.1.9 Data cleaning and coding

Sri Lanka

I excluded participants who had missing data for date of birth, sex, serum creatinine or fasting glucose. eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation [50]. Furthermore, as in chapter 5, I used the DEGREE diagnostic cut-offs for reduced kidney function which is an eGFR<60mL/min/1.73m in the absence of diabetes (fasting glucose >=7mg/l), hypertension (systolic ≥140 mm Hg, or diastolic ≥90 mm Hg), and heavy proteinuria (ACR>=30mg/mmol), based on a single measure [1].

In addition to the core DEGREE questionnaire (see chapter 5), an environmental questionnaire was administered during the interview, which collected data on lifestyle and environmental factors including use of pesticides, crops cultivated, chena cultivation (a high impact "slash and burn technique"),and drinking water sources used. This environmental questionnaire was adapted from a more detailed questionnaire which is described in chapter 6. I re-grouped co-variates with multiple categories into broader groups. Income and school years were grouped into tertiles. As the sample population are entirely rural, and most likely to conduct subsistence farming practices outside of their official employment (see appendix 5 table 1 for original categories), I re-grouped this category into "agricultural" and "non-agricultural" occupation. Smoking and alcohol consumption were combined to form a broad lifestyle covariate "non-drinker and/or smoker" and "drinker and/or smoker". Water source data was re-grouped into a single variable with four sub-categories "deep well", "shallow well", "tube well" and "other" source.

India

Previous analysis has been conducted using this dataset as part of an exploratory investigation into the presence of CKDu in India [5]. The findings from that research have resulted in India conducting a standardized DEGREE protocol study which is currently underway, but the findings are not yet available.

Participants with missing variables including serum creatinine, age and sex were excluded, as were diabetic (≥7mmol/L), hypertensive (systolic blood pressure ≥140mm Hg, or diastolic blood pressure ≥90mm Hg) and proteinuric (Albumin/Creatinine ratio (ACR) in urine ≥30 mg/mmol) participants. eGFR was calculated using the CKD-EPI equation. Detailed information on data collection methods in the field are described elsewhere[5]. I was sent participant coordinates by the PHFI (decimal degrees format) in a separate file which I merged with the harmonized CKDu data set using the participant study number as the common identifier. Finally, I re-grouped the variables as described for the Sri Lanka data. For original variables see appendix 6 table 1.

4.3 Results

4.3.1 Sri Lanka

4.3.1.1 Study population characteristics

The study population comprised 3296 adult participants over 18 years (males = 1016; females = 2280) who were non-diabetic (fasting glucose <7mg/l), non-hypertensive (systolic <140 mm Hg, or diastolic<90 mm Hg) and non-proteinuric (ACR<30mg/mmol) as per DEGREE diagnostic cut-offs [1]. Figure 27 shows the final sample size after exclusions.

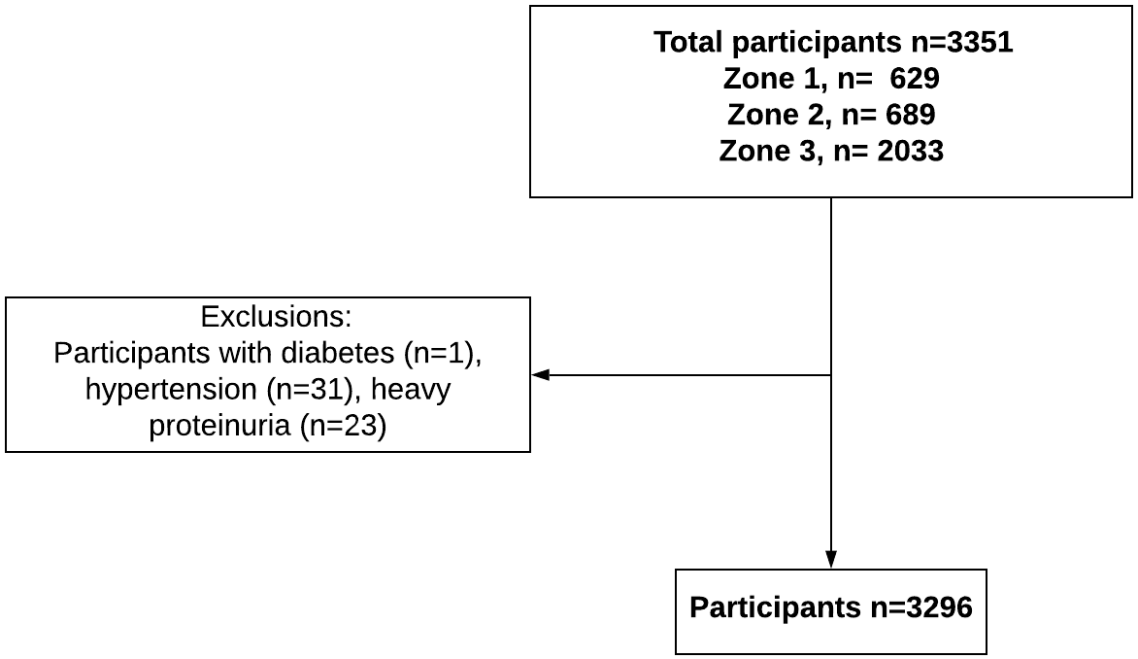


Figure 27. Study flowchart with exclusion criteria for the Sri Lanka DEGREE population sample

The mean (standard deviation (±SD)) participant age was 42±13.5 years. Mean BMI was 22.7±4.8 kg/m^2, and mean FFM was 26.2±13.1 kg/m^2. Mean systolic and diastolic blood pressure was 113.7±11.5 mm Hg, and 70.6±7.9 mm Hg respectively. Mean fasting plasma glucose was 111.3±19.3 mmol/l and the median (interquartile range (IQR) albumin/creatinine ratio (ACR) was 3.77(2.0, 6.3) mg/mmol. Approximately 70% of the participants were conducting daily agricultural labour (females n=1473 [65%], males n=839 [83%]). Overall, 65% of the population had completed >10 years of formal education.

4.3.1.2 Environmental characteristics

The hold out validation statistics for the heat index model fit using the EBK interpolation method (seen in table 18) across Sri Lanka showed a relatively good fit (R^2= 0.94)) with a low error (RMSE= 0.064) of the interpolated heat index data. See appendix 7, figure 1 for validation model fit regression line.

Table 18. Empirical Bayesian Kriging cross-validation summary for heat index across Sri Lanka

Prediction errors	Sri Lanka
Mean absolute error	0.078
Root Mean Square Error	0.064
R^2	0.941

The interpolated maps of the study zone show an increasing heat index gradient from south to north. As expected with a small study area (50km), there was a very small heat index range of 32.2-33.5^{o}C across the three study zones (figure 28a) (the bilinear heat index range was 32.4-33.2 ^{o}C), which were situated at a relatively low altitude ranging between 78-149 metres, with most of the population (66%) residing at altitudes over 100m (figure 28c). This population was entirely rural with 37% living within 200m of "cropland", and 64% living within 200m of "tree cover" land categories (figure 28b). For the environmental model surfaces for Sri Lanka see figure 29a-c. For data distribution and model fit for sensitivity analyses please see appendix 8 figures 1-3).

For the kriged surface predicted standard error maps can be created (figure 30c). The error map (figure 30c) shows the distribution of error across the heat index surface, where the areas of highest uncertainty are seen in the Central province of Sri Lanka, at the highest altitudes (2000-2500m). Issues with prediction error when interpolating temperature at high altitudes feature in the literature, and there are some potential reasons put forward, most prominent of which is the sparsity of monitoring stations at higher altitudes which could increase uncertainty of model outputs [254, 255]. ERA5 uses weather observations where such observations are available for modelling data. On top of these observations ERA5 uses a weather forecasting model to produce spatially and temporally continuous data, and the resulting data therefore contains some uncertainty. The highest error was observed between 2000-2500m (1.01-1.18 ^{o}C error) (figure 30c) which could be due to the sparsity of monitoring

data at higher altitudes which have heterogenous surfaces and are therefore more complex to model. Across the study zones in question however, the error was very low (0.037-0.039 ^{o}C uncertainty). Although this level of uncertainty was not seen in the DEGREE study zones, this finding is interesting and requires more complex future investigation to elucidate why this pattern is being observed.

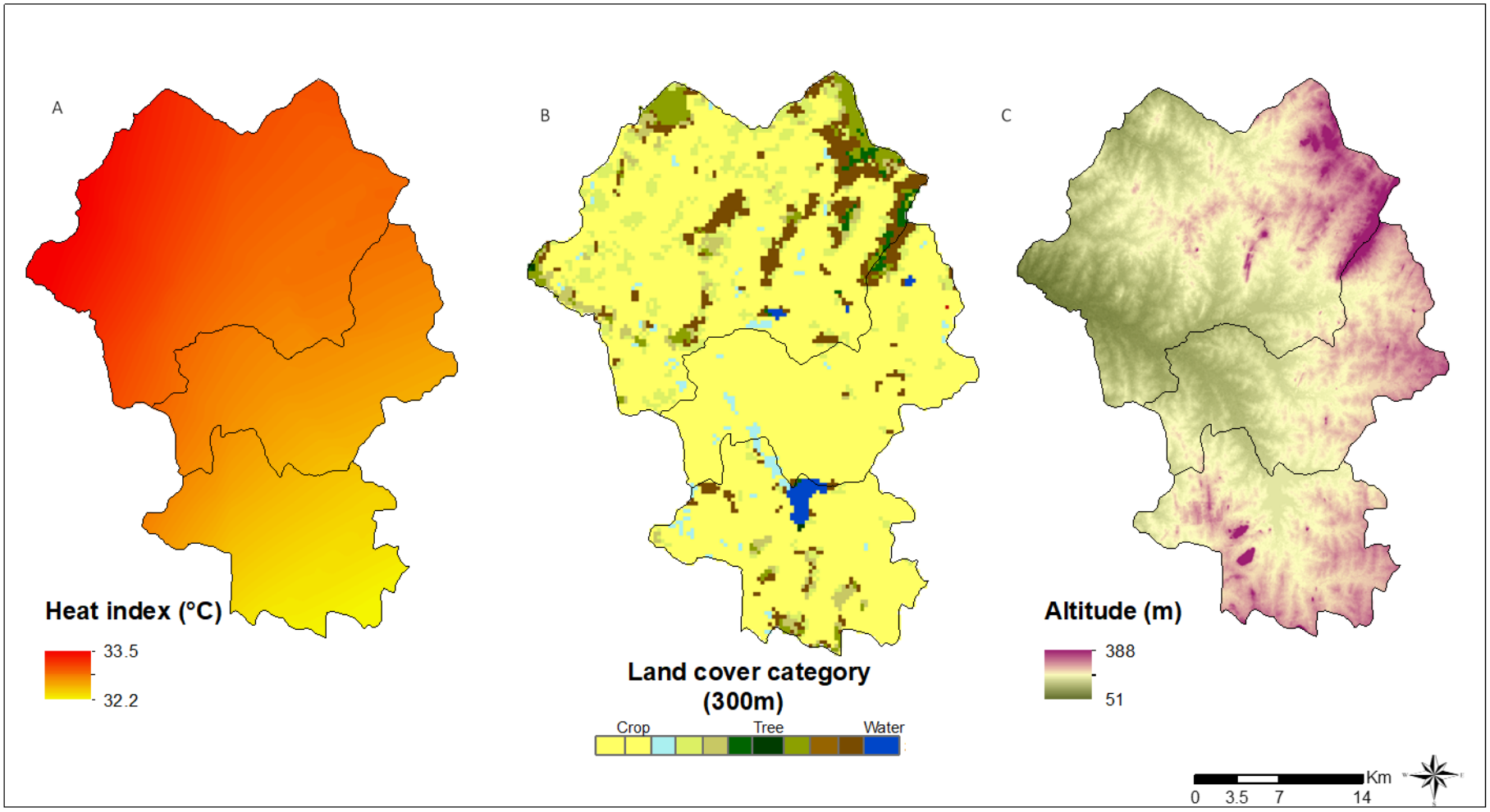


Figure 28. Environmental variable surfaces across Sri Lanka DEGREE study zones showing (a) Heat index, (b) Land cover (see appendix 4 for entire legend), (c) Altitude

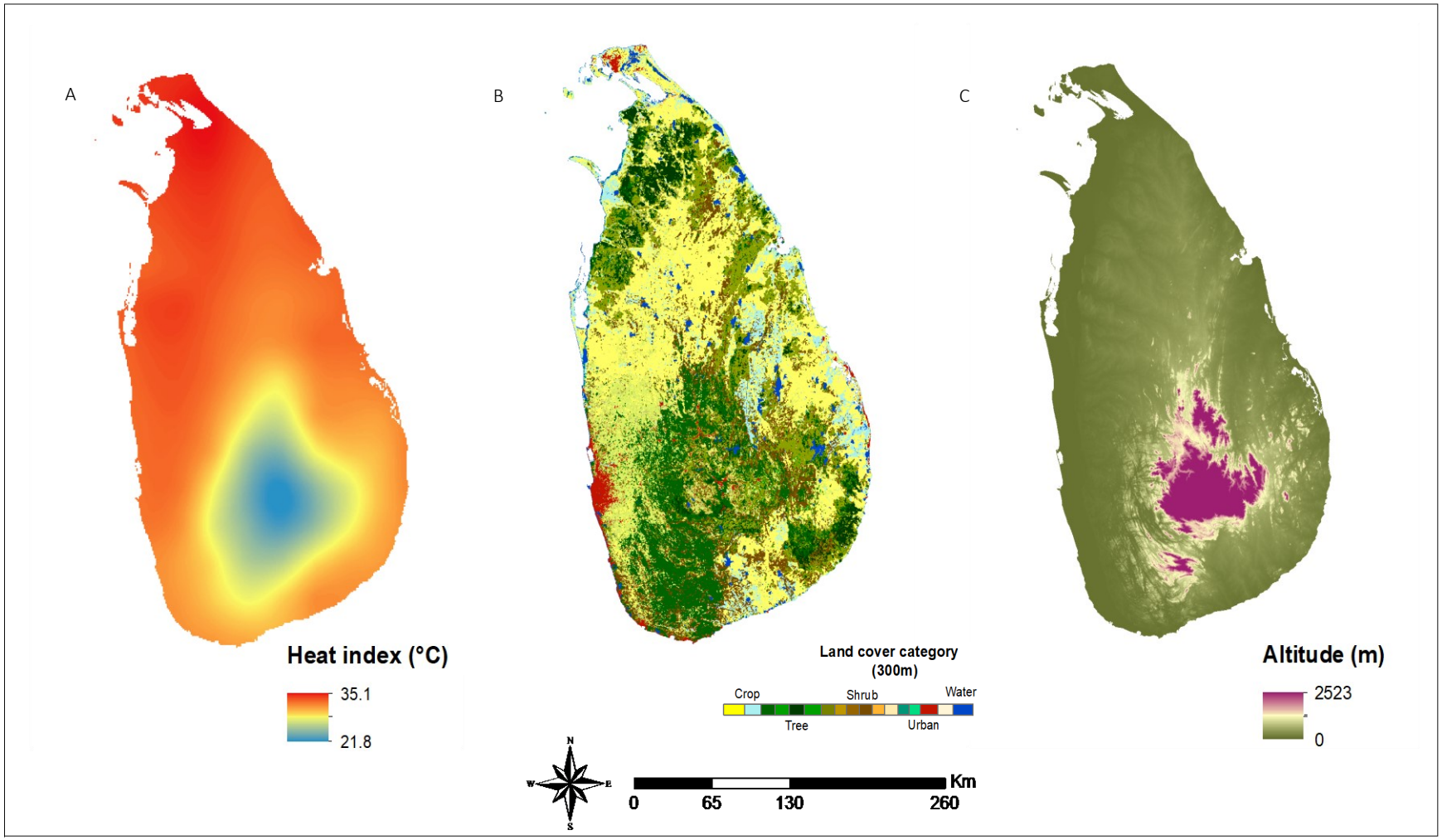


Figure 29. Environmental variable surfaces across Sri Lanka showing (a) Heat index, (b) Land cover (see appendix 4 for entire legend), (c) Altitude

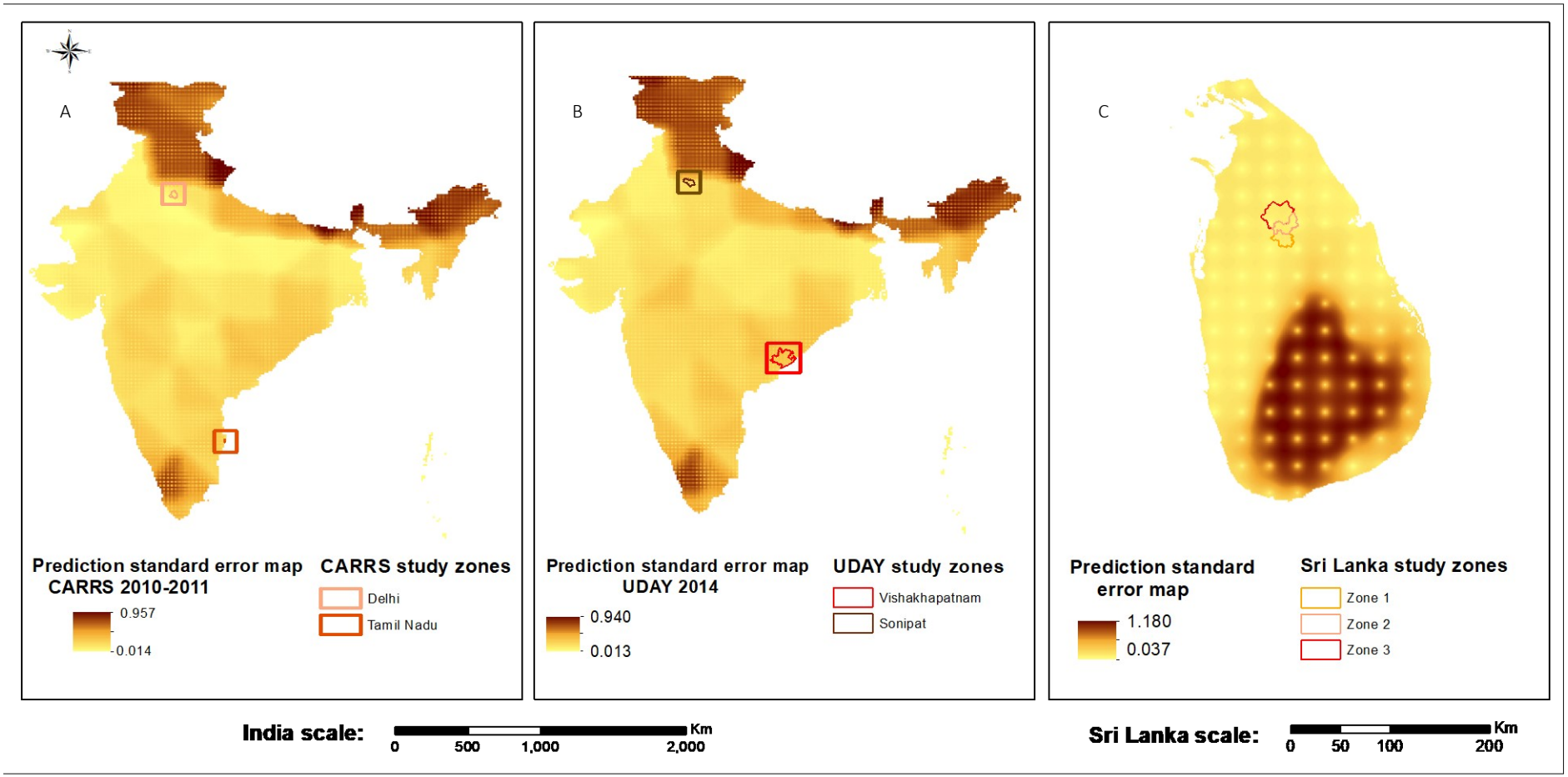


Figure 30. Prediction standard error maps for Empirical Bayesian Kriging (EBK) interpolation of heat index data across (a) India for CARRS study period (b) India for UDAY study period (c) Sri Lanka

Sensitivity analysis

I conducted a sensitivity analysis in order to understand whether the EBK interpolation method (scaled down from the national to study zone level) was indeed "fit for purpose" as a proposed interpolation method for the ERA-5 dataset. Table 19 shows the comparative results from the validation statistics for IDW, kriging, bilinear interpolation and the initial EBK model.

Results show that overall bilinear interpolation was the most accurate method in comparison to IDW and kriging. For bilinear interpolation, the RMSE (0.020) and MAE (0.017) had the overall lowest values which would suggest this was a better fit, and most appropriate method to use for the ERA-5 data across Sri Lanka (see appendix 8, figures 1-3 for data distribution and validation model fit regression lines).

Table 19. K-fold cross validation Summary of the performance of interpolation methods to predict heat index across Sri Lanka

Interpolation method	Efficiency	Error	
	R^2	RMSE	MAE
IDW	0.978	0.031	0.028
Kriging	0.973	0.131	0.023
Bilinear interpolation	0.982	0.020	0.017
EBK	0.941	0.064	0.078

R^2 = coefficient of determination, RMSE = root mean square error, ME = mean error, MRE = mean relative error, IDW: inverse distance weighting; EBK: Empirical Bayesian Kriging

In terms of Pearson correlation, the results of, Kriging, IDW, bilinear interpolation and EBK are significantly correlated with the original ERA-5 modelled data at 0.01 level (two-tailed), of which the correlation coefficients for the bilinear interpolation, IDW interpolation and the original ERA-5 modelled data were the strongest at 0.96 respectively.

The correlation coefficient for the initial EBK model used (10-15 neighbours) was lower than those for the more simplistic interpolation methods and was significant at the 0.05 level. This would suggest that the simpler interpolation approach would be most appropriate for this

dataset. Overall, among the interpolation methods, the performance of bilinear interpolation was best in comparison with other interpolation models

Table 20. Correlation analysis of interpolation methods across Sri Lanka

Sri Lanka	Original HI data	Kriging	Bilinear	IDW	EBK 10-15 neighbours
Pearson correlation	1.00	0.94	0.96	0.96	0.83
p	0.000	0.000	0.000	0.00	0.05

As a result of the sensitivity analyses showing that in comparison to EBK and other interpolation methods, I conducted regression analyses using EBK heat index data and bilinear interpolation data to assess whether these differing methods (with differing errors) would indeed impact the results across Sri Lanka.

I therefore present the main linear and logistic regression models simultaneously with EBK data and bilinear interpolation data. In models where the coefficients remain unchanged with bilinear heat index data, I append the results at the bottom of the table.

4.3.1.3 Mean eGFR and prevalence of low eGFR (<60ml/min/1.73m^2)

The mean eGFR was 93.9±23.4 mL/min/1.73m^2. An inverse relationship was observed between increasing age, fat free mass, heat index and altitude. Participants residing in the northernmost study zone (study zone 3), working in agriculture, cultivating crops, and those who were smokers and/or alcohol drinkers had a lower eGFR. Overall, males had a lower mean eGFR (90.1±26.6) than females (95.7±21.7). Figure 31 shows a notched box plot of eGFR by sex and study zone.



Figure 31. Distribution of estimated glomerular filtration rate (eGFR) levels by sex and study zone in Sri Lanka

The prevalence of eGFR<60 was 5.8% (95% confidence interval (95%CI) 5.1-6.7). In corroboration with the governmental CKD/u endemicity classification[6], the highest prevalence of eGFR<60 was observed in zone 3 (6.6% (95%CI 5.6-7.8) and the lowest prevalence was observed in zone 1 (3.5% (95%CI) 2.2-5.3). Zone 2 had a prevalence of 5.7% (95%CI 4.1-7.7). A significant difference in sex-specific prevalence of eGFR<60 was observed, with 11% (95%CI 9.0-13.0) of males versus 3.6% (95%CI 2.9-4.4) of females affected.

4.3.1.4 Risk factors for reduced eGFR and eGFR<60 using heat index derived from Empirical Bayesian kriging

Tables21 and 22 show linear and logistic regression models which are mutually adjusted for age and sex (minimally adjusted models 1), and are incrementally modelled with each environmental variable (models 3-5), and finally are fully adjusted for all socioeconomic and environmental risk factor variables (tables 21 and 22 model 6 respectively). I modelled heat index and altitude separately, and I present results for the heat index model with altitude coefficients appended.

Age was a key risk factor for reduced eGFR, with a decrease of 9.92 ml/min/1.73m^2 (-10.51, -9.33) per 10-year age increase (model 6). Additionally, increasing altitude and heat index (-5.91(95%CI) =-8.96, -2.87), and (-3.89(95%CI) =-4.36, -1.43) respectively), vegetable farming

and drinking and/or smoking were associated with a lower eGFR. There was no association observed between low eGFR and land cover (table 21 model 6). A protective effect was observed amongst males (7.78 (95%CI) = 5.24, 10.33).

Similar to the linear regression model, the odds of eGFR<60 (table 22 model 6) increased with age [OR per 10 year age increase (95%CI) = 2.46 (2.11, 2.88)], and smokers and/or drinking [OR(95%CI) = 2.43 (1.28, 4.41)]. There was also an increased risk of eGFR<60 amongst chena cultivators [OR (95%CI) = 1.47 (0.90, 2.51)]. There was no association between sex or any other environmental variables and risk of eGFR<60 in the fully adjusted logistic model (table 22, model 6).

Table 21. Associations of sociodemographic and anthropometric characteristics with low eGFR in participants without diabetes, hypertension and heavy proteinuria using Empirical Bayesian kriging derived heat index data in Sri Lanka, n=3297

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]
			eGFR coefficient (95% CI)			
Age [g]	-10.78***	-9.95***	-9.95	-9.84***	-9.87***	-9.92***
(10-year increase)	(-11.26, -10.34)	(-10.47, -9.29)	(-10.55, -9.37)	(-10.43, -9.26)	(-10.46, -9.29)	(-10.51, -9.33)
Sex [h]						
Male	-1.16	6.64***	7.81***	5.63***	5.22***	7.78***
	(-2.53, 0.20)	(3.27, 7.86)	(5.26, 10.37)	(3.34, 7.93)	(2.93, 7.53)	(5.24, 10.33)
Female	REF	REF	REF	REF	REF	REF
Education (years)						
≤5	0.50	0.35	0.37	0.41	0.33	0.31
	(-1.49, 2.50)	(-1.57, 2.52)	(-1.67, 2.42)	(-1.63, 2.46)	(-1.71, 2.38)	(-1.72, 2.16)
>5≤10	-2.24**	-1.49	-1.50	-1.46	-1.72	-1.46
	(-4.88, -0.55)	(-3.24, 0.22)	(-3.23, 0.21)	(-3.19, 0.26)	(-3.45, 0.21)	(-3.45, 0.25)
>10	REF	REF	REF	REF	REF	REF
Occupation						
Agricultural worker	-3.34***	-1.36	-1.80	-1.62	-1.70	-1.70
	(-4.88, -1.80)	(-4.51, 1.04)	(-4.57, 0,97)	(-4.40, 1.14)	(-4.48, 1.06)	(-4.47, 1.06)
Non-agricultural worker	REF	REF	REF	REF	REF	REF
Pesticide user						
Yes	-2.91***	-1.07	-1.05	-1.03	-1.03	-0.95
	(-4.30, -1.52)	(-2.71, 0.44)	(-2.63, 0.52)	(-2.60, 0.54)	(-2.61, 0.54)	(-2.53, 0.61)
No	REF	REF	REF	REF	REF	REF
Household monthly income (RS)[i]						
≤30,000	-0.73	-0.51	-0.62	-0.53	-0.42	-0.64
	(-2.04, 0.57)	(-1.86, 0.83)	(-1.97, 0.72)	(-1.88, 0.80)	(-1.77, 0.91)	(-1.98, 0.70)
>30,000	REF	REF	REF	REF	REF	REF
Healthy lifestyle choices						
Non-drinker and/or smoker	REF	REF	REF	REF	REF	REF
Drinker and/or smoker	-8.49***	-7.99***	-7.22***	-8.10***	-7.75***	-7.37***
	(-10.88, -6.11)	(-10.40, -5.59)	(-9.66, -4.80)	(-10.51, -5.71)	(-10.16, -5.35)	(-9.80, -4.94)
BMI (kg/m^2)	-0.39	0.26	0.33	-0.23	-0.34	-0.34
5kg/m^2 increase	(-1.04, 0.25)	(-0.93, 0.40)	(-0.39, 1.06)	(-0.89, 0.43)	(-1.01, 0.32)	(-1.01, 1.06)
Fat Free Mass (kg)	-0.90***	-0.06	-0.75	-0.05	-0.06	-0.71

5 kg/m2 increase	(-1.23, -0.59)	(-0.18, -0.06)	(-1.12, 0.39)	(-0.17, 0.06)	(-0.19, 0.05)	(-1.09, -0.35)
Water source						
Deep well	-1.42	-1.66	-1.64	-1.68	-1.70	-1.66
	(-3.83, 0.98)	(-4.05, 0.73)	(-4.02, 0.74)	(-4.07, 0.70)	(-4.09, 0.68)	(-4.53, 1.44)
Shallow well	REF	REF	REF	REF	REF	REF
Tube well	-1.15	-1.44	-1.48	-1.51	-1.07	-1.54
	(-4.16, 1.85)	(-4.45, 1.55)	(-4.48, 1.51)	(-4.51, 1.48)	(-4.07, 1.93)	(-2.27, 1.44)
Other source	1.69	1.43	1.32	1.18	1.68	1.80
	(-1.34, 4.73)	(-1.58, 4.46)	(-1.68, 4.34)	(-1.83, 4.20)	(-1.33, 4.70)	(-1.92, 4.10)
Chena cultivation						
Yes	-2.43***	-0.92	-0.96	-1.13	-0.46	-1.16
	(-3.82, -1.06)	(-2.72, 0.88)	(-2.76, 0.83)	(-2.93, 0.66)	(-2.27, 1.35)	(-1.60, 1.22)
No	REF	REF	REF	REF	REF	REF
Rice farmer						
Yes	-2.21 **	0.79	0.95	0.61	0.66	076
	(-3.66, -0.76)	(-1.47, 3.07)	(-1.31, 3.23)	(-1.66, 2.88)	(-1.60, 2.93)	(-3.40, 3.04)
No	REF	REF	REF	REF	REF	REF
Vegetable farmer						
Yes	-2.93**	-1.86*	-1.88*	-1.86*	-1.76*	-1.88*
	(-4.47, -1.41)	(-3.49, -0.23)	(-3.51, -0.25)	(-3.49, -0.24)	(-3.40, -0.14)	(-3.51, -0.26)
No	REF	REF		REF	REF	REF
Land cover						
Cropland	0.99*		1.18			0.62
	(-0.29, 2.26)		(-0.09, 2.45)			(-0.67, 1.92)
Tree/herbaceous cover	REF					
Altitude						
50m increment	-3.03**				-5.84***	-5.91***
	(-4.47, -1.60)				(-8.89, -2.79)	(-8.96, -2.87)
Heat index EBK (ºC)						
0.5 ºC increment	-3.03**			-3.01***		-3.89***
	(-4.47, -1.60)			(-4.48, -1.55)		(-4.36, -1.43)

[a] minimal adjustment for age, sex, [b] base variables mutually adjusted; [c] Base variables adjusted for land cover [d] Base variables adjusted for heat index; [e] Base variables adjusted for heat altitude, [f] Fully adjusted model (base + environmental variables), [g] adjusted for sex, [h] adjusted for age, [i] Exchange rate (RS to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

Table 22. Associations of sociodemographic and anthropometric characteristics with eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in Sri Lanka modelled using Empirical Bayesian kriging derived heat index data, n=3297

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]
			eGFR <60 OR (95%CI)			
Age [g]	2.80***	2.46***	2.47***	2.45***	2.46***	2.46***
(10-year increase)	(2.47, 3.21)	(2.12 2.89)	(2.11 2.88)	(2.11 2.87)	(2.12 2.89)	(2.11 2.88)
Sex [h]						
Male	2.61***	1.41	1.34	1.43	1.41	1.44
	(1.90, 3.61)	(0.74 2.65)	(1.10 2.58)	(0.75 2.68)	(0.74 2.65)	(0.69 2.56)
Female	Ref	REF	REF	Ref	Ref	Ref
Education (years)						
≤5	1.51	1.35	1.36	1.37	1.35	1.38
	(0.97, 2.34)	(0.84 2.16)	(0.89 2.21)	(0.86 2.20)	(0.84 2.16)	(0.85 2.18)
>5≤10	1.58*	1.45	1.46	1.47	1.45	1.48
	(1.07, 2.36)	(0.96 2.21)	(1.02 2.25)	(0.98 2.24)	(0.96 2.21)	(0.68 2.21)
>10	REF	REF	REF	REF	REF	REF
Occupation						
Agricultural worker	1.34	1.10	1.10	1.10	1.10	1.10
	(0.78, 2.44)	(0.43 2.76)	(0.44 2.80)	(0.43 2.77)	(0.43 2.76)	(0.43 2.77)
Non-agricultural worker	REF	REF	REF	REF	REF	REF
Pesticide user						
Yes	0.99	0.81	0.81	0.81	0.81	0.81
	(0.69, 1.43)	(0.54 1.25)	(0.54 1.09)	(0.54 1.25)	(0.54 1.25)	(0.53 1.24)
No	REF	REF	REF	REF	REF	REF
Household monthly income (RS) [i]						
≤30,000	1.31	1.14	1.14	1.14	1.14	1.14
	(0.90, 1.94)	(0.77 1.72)	(0.61 1.73)	(0.77 1.72)	(0.77 1.72)	(0.77 1.73)
>30,000	REF	REF	REF	REF	REF	REF
Healthy lifestyle choices						
Non-drinker and/or smoker	REF	REF	REF	REF	REF	REF
Drinker and/or smoker	2.37**	2.46**	2.36**	2.43**	2.46**	2.43**
	(1.34, 4.33)	(1.37 4.59)	(1.37 4.42)	(1.35 4.53)	(1.37 4.59)	(1.28 4.41)
BMI (kg/m^2)	0.93	0.97	0.95	0.98	0.97	0.95
5kg/m^2 increase	(0.78, 1.11)	(0.80 1.17)	(0.82 1.20)	(0.81 1.18)	(0.80 1.17)	(0.77 1.16)
Fat Free Mass (kg)	1.04	1.00	1.02	1.00	1.00	1.02
5 kg/m2 increase	(0.97, 1.11)	(0.94 1.03)	(0.94 1.11)	(0.94 1.03)	(0.94 1.03)	(0.94 1.11)

Water source						
Deep well	1.46	1.49	1.49	1.51	1.49	1.51
	(0.78, 3.04)	(0.79 3.13)	(0.79 3.14)	(0.80 3.17)	(0.79 3.13)	(0.79 3.13)
Shallow well	REF	REF	REF	REF	REF	REF
Tube well	1.52	1.76	1.75	1.73	1.76	1.72
	(0.69, 3.55)	(0.79 4.12)	(0.78 4.01)	(0.78 4.04)	(0.79 4.12)	(0.79 4.11)
Other source	1.02	1.16	1.17	1.15	1.16	1.16
	(0.38, 2.71)	(0.43 3.11)	(0.43 3.14)	(0.42 3.09)	(0.43 3.11)	(0.43 3.15)
Chena cultivation						
Yes	1.49*	1.51*	1.51*	1.47*	1.51*	1.47*
	(1.01, 2.25)	(0.92 2.61)	(0.89, 2.49)	(0.90 2.52)	(0.92 2.61)	(0.91 2.91)
No	REF	REF	REF	REF	REF	REF
Rice farmer						
Yes	1.12	0.85	0.84	0.86	0.85	0.86
	(0.73, 1.77)	(0.47 1.69)	(0.46 1.78)	(0.47 1.72)	(0.47 1.69)	(0.46 1.68)
No	REF	REF	REF	REF	REF	REF
Vegetable farmer						
Yes	1.25	1.30	1.29	1.29	1.30	1.27
	(0.87, 1.78)	(0.88 1.90)	(0.88 1.89)	(0.88 1.89)	(0.88 1.90)	(0.88 1.90)
No	REF	REF	REF	REF	REF	REF
Land cover						
Cropland	0.79		0.77			0.79
	(0.57, 1.11)		(0.54 1.09)			(0.54 1.31)
Tree/herbaceous cover	REF		REF			REF
Altitude						
50m increment	1.15				0.69	0.69
	(0.79, 1.70)				(0.30 1.58)	(0.31 1.58)
Heat index EBK (°C)						
0.5 °C increment	0.82			1.10		1.09
	(0.38, 1.79)			(0.74 1.68)		(0.72 1.69)

[a] minimal adjustment for age, sex , [b] base variables mutually adjusted; [c] Base variables adjusted for land cover [d] Base variables adjusted for heat index; [e] Base variables adjusted for heat index, [f] Fully adjusted model (base + environmental variables), [g] adjusted for sex, [h] adjusted for age, [i]Exchange rate (RS to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

4.3.1.5 Risk factors for reduced eGFR and eGFR<60 using heat index derived from bilinear interpolation

In the linear and logistic regression models I created using the bilinear interpolation heat index data (tables 23 and 24), there were some marked changes in coefficients observed in comparison to those in tables 21 and 22, particularly amongst males.

In the fully adjusted linear regression model (6) in table 23, the protective effect observed amongst males was lost which could suggest that with the use of the more accurate predictive model for heat index prediction (and therefore assignment of values), this association is likely not as strong as initially predicted.

The heat index coefficient itself also had a more negative coefficient in the fully adjusted model (table 23 model 6) which suggests that in this setting heat index could be more of a risk factor than previously estimated in linear regression models.

Table 23. Associations of sociodemographic and anthropometric characteristics with low eGFR in participants without diabetes, hypertension and heavy proteinuria using bilinear interpolation derived heat index data in Sri Lanka, n=3297

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]
			eGFR			
			Coefficient (95%CI)			
Age [g]	-10.78***	-8.36***	-8.36***	-8.38***	-8.33***	-8.33***
(10-year increase)	(-11.26, -10.34)	(-8.99, -7.74)	(-9.00, -7.74)	(-9.02, -7.76)	(-8.96, -7.71)	(-8.97, -7.71)
Sex [h]						
Male	-1.16	0.82	1.79	0.97	0.94	2.46
	(-2.53, 0.20)	(-2.37, 0.72)	(-2.34, 2.75)	(-2.52, 0.57)	(-2.49, 0.60)	(-1.46, 10.63)
Female	REF	REF	REF	REF	REF	REF
Education (years)						
≤5	0.50	-1.37	-1.71	-1.37	-1.72	1.79
	(-1.49, 2.50)	(-6.63, -1.33)	(-9.41, 5.22)	(-8.83, 5.28)	(-9.38, 5.20)	(-9.31, 5.14)
>5≤10	-2.24**	-1.08*	-1.06	-1.08	-1.06	-2.64
	(-4.88, -0.55)	(-9.47, -1.01)	(-8.82, 5.32)	(-9.47, 2.57)	(-2.03, 3.15)	(-8.45, 5.10)
>10	REF	REF	REF	REF	REF	REF
Occupation						
Agricultural worker	-3.34***	-2.08	-3.07	-1.87	-1.89	-4.07
	(-4.88, -1.80)	(-12.03, -0.15)	(-12.02, -2.13)	(-11.92, -0.04)	(-12.03, -1.15)	(-8.47, 1.04)
Non-agricultural worker	REF	REF	REF	REF	REF	REF
Pesticide user						
Yes	-2.91***	-0.72	-0.70	-0.75	-0.66	-0.64
	(-4.30, -1.52)	(-2.41, 0.96)	(-2.39, 0.98)	(-2.44, 0.93)	(-2.35, 1.01)	(-2.33, 1.03)
No	REF	REF	REF	REF	REF	REF
Household monthly income (RS) [i]						
≤30,000	-0.73	-1.40	-1.41	-1.39	-1.58	-1.59
	(-2.04, 0.57)	(-2.93, 0.11)	(-2.94, 0.10)	(-2.92, 0.12)	(-3.10, 0.06)	(-3.11, 0.07)
>30,000	REF	REF	REF	REF	REF	REF
Healthy lifestyle choices						
Non-drinker and/or smoker	REF	REF	REF	REF	REF	REF
Drinker and/or smoker	-8.49***	-4.24***	-4.26***	-4.14***	-4.21***	-4.23***
	(-10.88, -6.11)	(-6.36, -2.13)	(-6.38, -2.15)	(-6.26, -2.03)	(-6.33, -2.10)	(-6.34, -2.12)

BMI (kg/m^2)	-0.39	-0.37	-0.37	-0.43	-0.45	-0.44
5kg/m^2 increase	(-1.04, 0.25)	(-1.08, 0.14)	(-1.01, 1.10)	(-1.17, 0.31)	(-1.19, 0.28)	(-1.19, 0.29)
Fat Free Mass (kg)	-0.90***	-0.04	-0.37	-0.04	-0.05	-0.06
5 kg/m2 increase	(-1.23, -0.59)	(-0.29, 0.39)	(-1.11, 0.37)	(-0.30, 0.39)	(-0.28, 0.40)	(-0.28, 0.41)
Water source						
Deep well	-1.42	-0.89	-0.84	-0.95	-0.88	-0.82
	(-3.83, 0.98)	(-1.77, 3.56)	(-0.29, 0.40)	(-1.71, 3.61)	(-1.77, 3.54)	(-1.83, 3.48)
Shallow well	REF	REF	REF	REF	REF	REF
Tube well	-1.15	-1.59	-2.54	-3.82	-3.61	-1.68
	(-4.16, 1.85)	(0.25, 6.94)	(-0.20, 6.89)	(0.48, 7.17)	(0.28, 6.94)	(-2.22, 6.89)
Other source	1.69	3.21	3.21	3.22	2.95	2.95
	(-1.34, 4.73)	(-0.26, 6.68)	(-0.25, 6.68)	(-0.24, 6.69)	(-0.51, 6.41)	(-0.50, 6.41)
Chena cultivation						
Yes	-2.43***	-0.87	-0.84	-0.46	-0.43	-1.41
	(-3.82, -1.06)	(-0.81, 2.56)	(-0.84, 2.53)	(-1.25, 2.17)	(-0.27, 3.14)	(-2.29, 3.12)
No	REF	REF	REF	REF	REF	REF
Rice farmer						
Yes	-2.21 **	0.33	0.33	0.39	0.01	0.01
	(-3.66, -0.76)	(-1.85, 2.51)	(-1.84, 2.52)	(-1.78, 2.57)	(-2.17, 2.19)	(-2.16, 2.20)
No	REF	REF	REF	REF	REF	REF
Vegetable farmer						
Yes	-2.93**	-2.14	-2.15	-2.07	-1.68	-1.68*
	(-4.47, -1.41)	(-2.63, 3.66)	(-2.64, 3.67)	(-2.56, 3.59)	(-0.15, 3.21)	(-4.51, -0.20)
No	REF	REF	REF	REF	REF	REF
Land cover						
Cropland	0.99*		-3.99			2.45
	(-0.29, 2.26)		(-12.08, 0.09)			(-2.52, 4.40)
Tree/herbaceous cover	REF					
Altitude						
50m increment	-3.03**			-4.80*		-4.93***
	(-4.47, -1.60)			(-8.28, -1.33)		(-8.41, -1.47)
Heat index (ºC)						
0.5 ºC increment	-6.20**				-4.36***	-4.40***
Bilinear interpolation	(-8.21, -4.21)				(-6.56, -2.16)	(-6.61, -2.21)

[a] minimal adjustment for age, sex , [b] base variables mutually adjusted; [c] Base variables adjusted for land cover [d] Base variables adjusted for heat index; [e] Base variables adjusted for heat index, [f] Fully adjusted model (base + environmental variables), [g] adjusted for sex, [h] adjusted for age, [i] Exchange rate (RS to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

Like the linear regression models, logistic regression models showed minimal changes in coefficients. A very minor increased risk was observed with increasing heat index [OR(95%CI) = 1.02 (0.72, 1.69)], (table 24 model 6) which suggests that in this setting heat index could be slightly more of a risk factor than previously estimated.

Furthermore, throughout all the models across tables 21-24 there were marked changes in coefficients between minimal and fully adjusted models, particularly between the work group variable and chena cultivation variable whereby the association between eGFR and work group was lost when adjusted for chena farming. I conducted a further sensitivity analysis to model the fully adjusted variables (in a linear regression model) again in the absence of chena cultivation and the association between agricultural labour and reduced eGFR was restored, which would suggest that the presence of both these variables in one model are causing spurious results. For the sensitivity analysis see appendix 9 table 1.

Table 24. Associations of sociodemographic and anthropometric characteristics with eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in Sri Lanka modelled using bilinear interpolation derived heat index data, n=3297

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model4[d]	Model 5[e]	Model 6[f]
			eGFR <60 OR (95%CI)			
Age [g]	2.80***	2.54***	2.32***	2.33***	2.34***	2.34***
(10-year increase)	(2.47, 3.21)	(1.97, 2.77)	(1.97, 2.77)	(1.97, 2.78)	(1.98, 2.80)	(1.98 2.80)
Sex [h]						
Male	2.61***	1.91	2.06	2.05	2.17	2.17
	(1.90, 3.61)	(1.38, 3.11)	(1.38, 3.11)	(1.37, 3.09)	(1.45, 3.29)	(1.45 3.29)
Female	REF	REF	REF	REF	REF	REF
Education (years)						
≤5	1.51	2.98	2.55	2.55	2.61	2.60
	(0.97, 2.34)	(1.59, 4.14)	(1.58, 4.12)	(1.58, 4.13)	(1.62, 4.24)	(1.61 4.21)
>5≤10	1.58*	2.73	2.34	2.34	2.25	2.25
	(1.07, 2.36)	(1.54, 3.60)	(1.54, 3.60)	(1.54, 3.60)	(1.47, 3.47)	(0.79 3.13)
>10	REF	REF	REF	REF	REF	REF
Occupation						
Agricultural worker	1.34	1.10	1,11	1.16	1.11	1.11
	(0.78, 2.44)	(1.65, 7.69)	(1.74, 6.02)	(1.06, 4.15)	(1.10, 3.57)	(0.11 2.53)
Non-agricultural worker	REF	REF	REF	REF	REF	REF
Pesticide user						
Yes	0.99	0.78	0.78	0.72	0.77	0.77
	(0.69, 1.43)	(0.51, 1.90)	(0.51, 1.19)	(0.51, 1.18)	(0.51, 3.29)	(0.51 1.18)
No	REF	REF	REF	REF	REF	REF
Household monthly income (RS) [i]						
≤30,000	1.31	1.30	1.28	1.27	1.29	1.30
	(0.90, 1.94)	(0.83, 2.02)	(0.83, 2.03)	(0.82, 1.01)	(0.84, 2.04)	(0.84 2.05)
>30,000	REF	REF	REF	REF	REF	REF
Healthy lifestyle choices						
Non-drinker and/or smoker	REF	REF	REF	REF	REF	REF
Drinker and/or smoker	2.37**	2.67***	2.78	2.84***	2.68***	2.67**
	(1.34, 4.33)	(1.69, 4.68)	(1.69, 4.66)	(1.73, 4.76)	(1.62, .450)	(1.62 4.49)
BMI (kg/m^2)	0.93	0.95	0.95	0.94	0.96	0.97
5kg/m^2 increase	(0.78, 1.11)	(0.78, 1.17)	(0.78, 1.17)	(0.77, 1.16)	(0.879, 1.19)	(0.79 1.19)

Fat Free Mass (kg)	1.04	0.98	0.98	0.98	0.98	0.98
5 kg/m2 increase	(0.97, 1.11)	(0.90, 1.07)	(0.90, 1.07)	(0.90, 1.07)	(0.90, 1.07)	(0.90 1.07)
Water source						
Deep well	1.46	1.29	1.29	1.28	1.23	1.23
	(0.78, 3.04)	(0.64, 2.83)	(0.65, 2.85)	(0.64, 2.81)	(0.61, 2.70)	(0.62 2.71)
Shallow well	REF					REF
Tube well	1.52	1.00	1.00	1.00	0.93	0.93
	(0.69, 3.55)	(0.41, 2.51)	(0.41, 2.51)	(0.42, 2.53)	(0.38, 2.34)	(0.38 2.34)
Other source	1.02	1.36	1.36	1.38	1.47	1.46
	(0.38, 2.71)	(0.48, 3.87)	(0.48, 3.87)	(0.48, 3.92)	(0.52, 4.14)	(0.52 4.13)
Chena cultivation						
Yes	1.49*	1.76*	1.76*	1.71*	1.72*	1.01*
	(1.01, 2.25)	(0.43, 2.29)	(0.54, 2.28)	(0.40, 2.22)	(0.41, 2.22)	(0.41 1.22)
No	REF	REF	REF	REF	REF	REF
Rice farmer						
Yes	1.12	1.12	1.18	1.12	1.19	1.18
	(0.73, 1.77)	(0.56, 2.09)	(0.56, 2.09)	(0.57, 2.11)	(0.59, 2.23)	(0.59 2.20)
No	REF	REF	REF	REF	REF	REF
Vegetable farmer						
Yes	1.25	0.80	0.83	0.83	0.90	0.92
	(0.87, 1.78)	(0.56, 1.23)	(0.57, 1.24)	(0.56, 1.23)	(0.61, 1.35)	(0.61 1.35)
No	REF	REF	REF	REF	REF	REF
Land cover						
Cropland	0.79		1.23			1.15
	(0.57, 1.11)		(0.05, 2.31)			(0.10 1.32)
Tree/herbaceous cover	REF		REF			REF
Altitude				0.53		
50m increment	1.15			(0.21, 1.33)		0.53
	(0.79, 1.70)					(0.23, 1.07)
Heat index bilinear interpolation (ºC)						
0.5 ºC increment	0.82				1.38	1.02*
	(0.38, 1.79)				(1.29, 4.53)	(0.72 1.69)

[a] minimal adjustment for age, sex , [b] base variables mutually adjusted; [c] Base variables adjusted for land cover [d] Base variables adjusted for heat index; [e] Base variables adjusted for heat index, [f] Fully adjusted model (base + environmental variables), [g] adjusted for sex, [h] adjusted for age, [i]Exchange rate (RS to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

4.3.1.6 Reduced eGFR and eGFR<60 by study site

As the prevalence of eGFR<60 varied widely across zones 1-3, I explored risk factors for reduced eGFR and eGFR<60 for each zone. I included variables associated with low eGFR and risk of eGFR<60 in the previous regression models (tables 21-24, model 6), and all environmental variables as these are of a priori interest.

Across zones 1-3 increasing age, heat index, altitude and drinking and/or smoking was associated with reduced eGFR (table 25, models 1a, 2a, 3a). A protective effect was observed amongst males across all zones. Chena cultivation was associated with a low eGFR in zone 1, and zone 2 deep well water drinkers also had a lower eGFR.

In logistic regression models, age remained a key risk factor for eGFR<60 across all zones (table 25, models 1b, 2b, 3b). In zones 1 and 2 there were no additional risk factors for eGFR<60 observed. In zone 3 drinking and/or smoking remained associate with eGFR<60, in addition to tube well drinking water drinkers and conducting chena cultivation practices ([OR (95%) = 2.81 (1.05, 8.51)], and [OR](95%)CI = 2.09 (1.15,3.98)] respectively).

As part of the sensitivity analyses for bilinear interpolation I modelled each study zone with bilinear interpolation data. Overall, the other coefficients were unchanged when modelled with the bilinear heat index data therefore I have appended the bilinear coefficients to table 25.

What was noted however is the negative association with heat index and eGFR in zones 1 and 2 (models 1a and 2a) which could suggest that the more accurate heat index data could be showing that increasing ambient temperatures could be more of a risk factor than previously observed using the EBK modelling in these study zones.

Table 25. Fully adjusted logistic regression by study area in Sri Lanka (n=3297)

Variable	Area 1		Area 2		Area 3	
	eGFR fully adjusted[a]	eGFR <60 fully adjusted[b]	eGFR fully adjusted[a]	eGFR <60 fully adjusted[b]	eGFR fully adjusted[a]	fully adjusted eGFR <60[b]
Age [c]	-9.51***	3.78***	-10.14***	2.34***	-10.20***	2.79***
(10-year increase)	(-10.52, -8.51)	(2.44, 6.46)	(-11.40, -8.89)	(1.78 3.15)	(-10.90, -9.51)	(2.34, 3.37)
Sex						
Male	8.60***	0.94	8.38***	1.30	3.59*	1.46
	(4.17, 13.05)	(0.19, 4.00)	(3.65, 13.12)	(0.39 3.9)	(0.48, 6.71)	(0.53, 3.69)
Female	REF	REF	REF	REF	REF	REF
Pesticide user						
Yes	-0.55	0.66	-1.03	1.05	-1.77	0.71
	(-3.74, 2.63)	(0.20, 2.16)	(-4.80, 2.72)	(0.43 2.6)	(-3.76 0.22)	(0.45, 1.31)
No	REF	REF	REF	REF	REF	REF
Healthy lifestyle choices						
Non-drinker and/or smoker	REF	REF	REF	REF	REF	REF
Drinker and/or smoker	-9.14***	2.18	-7.69**	1.63	-8.37***	3.31***
	(-13.61, -4.68)	(0.51, 1.05)	(-12.93, -2.45)	(0.53 5.43)	(-11.71, -5.03)	(1.37, 8.87)
Land cover						
Cropland	1.62	1.44	1.67	0.70	0.09	0.75
	(-1.18, 4.43)	(0.51, 4.04)	(-1.46, 4.81)	(0.31 1.52)	(-1.51, 1.70)	(0.48, 1.16)
Tree/herbaceous cover	REF	REF	REF	REF	REF	REF
Water source						
Deep well	0.09	0.51	-1.22	1.84	-2.95*	1.87
	(-5.39, 5.58)	(0.11, 3.19)	(-7.37, 4.92)	(0.48 1.24)	(-5.83, -0.08)	(0.83, 5.07)
Shallow well	REF	REF	REF	REF	REF	REF
Tube well	0.20	0.54	3.12	0.28	-3.42	2.81*
	(-5.98, 6.39)	(0.07, 4.43)	(-5.66, 11.91)	(0.00 9.76)	(-7.08, 0.23)	(1.05, 8.51)
Unknown	3.46	0.44	-4.94	0.92	-2.40	3.51
	(-2.36, 9.29)	(0.13, 2.74)	(-12.79, 2.91)	(0.04 1.11)	(-6.81, 1.97)	(0.81, 14.00)
Chena cultivation						
Yes	-5.07*	2.36	-1.78	1.53	-2.01	2.09*
	(-9.36, -0.78)	(0.43, 1.68)	(-5.48, 1.91)	(0.68 3.55)	(-4.10, 0.07)	(1.15, 3.98)
No	REF	REF	REF	REF	REF	REF

Rice farmer						
Yes	0.67	0.69	-1.49	0.73	0.63	0.74
	(-3.17, 4.53)	(0.16, 3.44)	(-5.38, 2.39)	(0.28 1.95)	(-1.61, 1.97)	(0.37, 1.67)
No	REF	REF	REF	REF	REF	REF
Vegetable farmer						
Yes	-2.50	1.08	-0.68	1.31	-1.41	1.18
	(-6.55, 1.54)	(0.21, 4.20)	(-4.93, 2.56)	(0.52 3.05)	(-3.31, 0.48)	(0.75, 1.86)
No	REF	REF	REF	REF	REF	REF
EBK Heat index (°C)	-3.95**	0.94	-9.86*	1.48	-3.16**	1.07
0.5°C increment	(-6.73, -1.18)	(0.33, 2.82)	(-19.04, -0.70)	(0.23 1.28)	(-4.94, -1.40)	(0.66, 1.78)
Altitude	-16.44**	1.59	-19.20*	3.50	-7.76**	1.80
50m increment	(-27.80, -5.09)	(0.21, 20.65)	(-29.05, -0.83)	(0.41, 3.89)	(-12.38, -2.98)	(0.34, 3.99)
Bilinear interpolation heat index data (0.5 °C)	-5.16***	0.14	-12.66*	0.18	-2.15**	1.65
	(-13.81, -4.41)	(0.01, 1.47)	(-17.46, -2.89)	(0.02, 1.78)	(-5.31, -1.08)	(0.33, 5.62)

[a] minimal adjustment for age, sex and area, [b] all variables mutually adjusted; [c] adjusted for sex and area

In multiple regression models I included risk factors which were associated with low eGFR and risk of eGFR <60, in the initial models, as well as environmental variables which are of a priori interest. In linear regression models, age, drinking and/or smoking were associated with a low eGFR (table 26, model 1).

Environmental variables associated with reduced eGFR were increasing heat index and altitude [regression coefficient (95%CI) = -3.06 (-4.53, -1.61) and regression coefficient (95%CI) = -5.52 (-8.55, -2.50) respectively], vegetable farming [regression coefficient (95%CI) = -2.21 (-3.79, -0.64)], and chena cultivation practices [regression coefficient(95%)CI= -1.70(-3.35,-0.06)]. Risk of eGFR<60 remained associated with age, drinking, and/or smoking and chena cultivation. No increased risk of eGFR<60 was observed for any other environmental risk factors (table 26 model 2).

Models using the bilinear heat index have very similar associations a seen in models 1 and 2, however the male coefficients in linear regression models were markedly lower than that observed in the model 1 (table 26). The heat index coefficients were also different, with the bilinear heat index data having a more negative coefficient [regression coefficient(95%CI) = -4.20, (-6.45, -1.97)] with eGFR than that observed with the EBK heat data.

Table 26. Multiple regression Sri Lanka heat index analyses comparing EBK method and bilinear interpolation data with altitude coefficients appended (n=3297)

Variable	Model 1 Fully adjusted	Model 2 Fully adjusted	Model 3 Fully adjusted	Model 4 fully adjusted
	eGFR Coefficient (95%CI)[a];	eGFR <60 OR (95%CI)[a];	eGFR Coefficient (95%CI)[b];	eGFR <60 OR (95%CI)[b];
Age[c]	-9.93***	2.46***	-8.93***	2.57***
(10-year increase)	(-10.51, -9.37)	(2.12, 2.88)	(-9.54, -8.33)	(2.19, 3.04)
Sex[d]				
Male	5.72***	1.44	3.38*	2.01
	(3.49, 7.97)	(0.77, 2.66)	(1.91, 1.15)	(1.36, 2.97)
Female	REF	REF	REF	REF
School years				
0-5	0.39	1.38	0.82	1.10
	(-1.59, 2.39)	(0.87, 2.21)	(-1.40, 1.69)	(0.18, 4.80)
5-10	-1.38	1.48	-1.86	0.95
	(-3.07, 0.31)	(0.98, 2.24)	(-2.02, -0.21)	(0.71, 4.00)
>10	REF	REF	REF	REF
Pesticide user				
Yes	-1.13	0.82	-4.23	1.06
	(-2.66, 0.40)	(0.54, 1.25)	(-5.79, -2.68)	(0.70, 1.61)
No	REF	REF	REF	REF
Healthy lifestyle choices				
Non-drinker and/or smoker	REF	REF	REF	REF
Drinker and/or smoker	-8.35***	2.43***	-5.15***	2.62***
	(-10.74, -5.96)	(1.35, 4.52)	(-5.93, -2.39)	(1.74, 3.98)
Heat index EBK (°C)	-3.06***	1.11	-	-
0.5 increments	(-4.53, -1.61)	(0.74, 1.68)		
Heat index bilinear interpolation (°C)	-	-	-4.20*	2.42
0.5 increments			(-6.45, -1.97)	(1.33, 4.63)
Land cover				
Cropland	0.49	0.79	-1.49	2.29

	(-0.79, 1.79)	(0.55, 1.13)	(-3.16, 0.56)	(0.01, 2.99)
Tree cover	REF	REF	REF	REF
Chena cultivation				
Yes	-1.70*	1.51*	-1.05*	1.04*
	(-3.35, -0.06)	(0.95, 2.45)	(-2.05, -0.03)	(0.37, 1.09)
No	REF	REF	REF	REF
Vegetable farmer				
Yes	-2.21**	1.30	1.57*	0.89
	(-3.79, -0.64)	(0.88, 1.90)	(0.02, 3.14)	(0.61, 1.34)
No	REF	REF	REF	REF
Rice farmer				
Yes	-0.18	0.90	-0.44	1.24
	(-1.93, 1.55)	(0.53, 1.57)	(-2.27, 2.18)	(0.63, 2.31)
No	REF	REF	REF	REF
Water source				
Deep well	-1.24	1.52	0.64	1.24
	(-4.19, 0.56)	(0.80, 3.17)	(-2.06, 3.36)	(0.62, 2.72)
Shallow well	REF	REF	REF	REF
Tube well	-1.43	1.73	-2.98	0.93
	(-4.41, 1.54)	(0.78, 4.04)	(-3.42, 0.39)	(0.38, 3.98)
Unknown	1.24	1.15	4.33*	1.14
	(-1.76, 4.26)	(0.42, 3.07)	(0.81, 7.86)	(0.41, 3.17)
Altitude (m)	-5.52***	0.69	-5.61***	0.65
(50m increments)	(-8.55, -2.50)	(0.31, 1.58)	(-9.15, -2.08)	(0.27, 1.59)

[a] Models mutually adjusted for age, sex, school years, use of pesticides, healthy lifestyle choice, land cover, chena, rice, vegetable cultivation, water source, altitude and heat index using the EBK methodology; [a] Models mutually adjusted for age, sex, school years, use of pesticides, healthy lifestyle choice, land cover, chena, rice, vegetable cultivation, water source, altitude and heat index using the bilinear interpolation methodology; [c] adjusted for sex; [d] adjusted for age

4.3.1.7 Linear mixed modelling

The caterpillar plot below (figure 32) shows the linear mixed model for Sri Lanka which was modelled using environmental variables (both satellite-derived and self-reported) as the fixed effects, and study zones as the random structures. The intercept denotes the overall mean eGFR across the study zones, and the blue circles represent the average eGFR for each zone. The black bars represent 95% confidence intervals. Results indicate that Zone 3 has a significantly lower eGFR than zones 1 and 2 when accounting for environmental variables. Significant negative effects on eGFR were observed with increasing heat index (-4.13 (-10.93, -3.52)), increasing altitude (-10.09 (-15.07, -5.11)), use of pesticides (-2.16 (-3.89, -0.44)), chena cultivation (-8.90 (-10.83, -6.97)) and rice cultivation (-7.87 (-9.89, -5.86)). Drinking water from sources other than groundwater had a positive association with eGFR (4.80 (1.12, 8.51)) (table 27). There was no significant difference observed when the data was modelled using bilinear interpolation heat index data (appended on table 27).

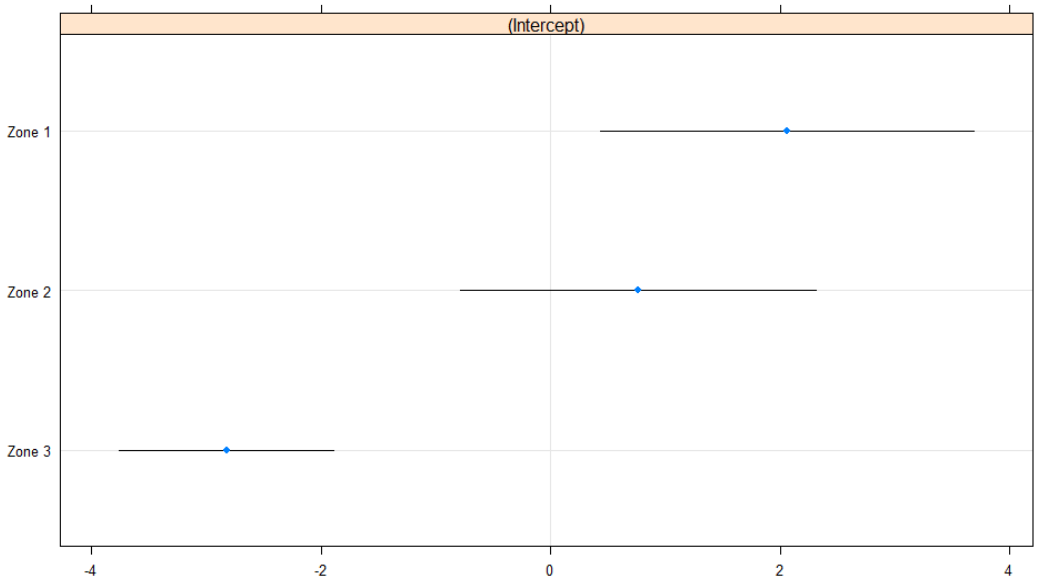


Figure 32. Linear mixed model caterpillar plot of eGFR in study zones 1-3 in Sri Lanka. Intercept denotes overall mean eGFR across the study zones; blue circles represent the average eGFR for each zone; black bars represent 95% confidence intervals.

Table 27. Fixed effects table of environmental factor effect (heat index, altitude, land cover, pesticide use, chena cultivation, rice and vegetable farming and water source) on eGFR across zones 1-3.

Variable	Regression coefficient	Confidence interval (95%)	P value
Heat index (EBK) (°C)	-4.13	-10.93, -3.52	<0.001
Land cover (crop)	1.88	-0.54, 2.57	0.20
Use of pesticides (yes)	-2.16	-3.89, -0.44	<0.001
Water source: other	4.80	1.12, 8.51	0.01
Deep well	1.06	-4.40, 1.35	0.29
Tube well	-0.26	-3.85, 3.31	0.51
Chena cultivation	-8.90	-10.83, -6.97	<0.001
Rice cultivation	-7.87	-9.89, -5.86	<0.001
Vegetable cultivation	1.71	-3.62, 0.19	0.07
Altitude (m)	-10.09	-15.07, -5.11	<0.001
Heat index (bilinear interpolation) (°C)	-4.91	-7.70, -2.12	<0.001

4.3.2 India

4.3.2.1 Study population characteristics

The study population comprised 11,119 adult participants over 18 years (males =4696 ; females =6423) who were non-diabetic (fasting glucose <7mg/l), non-hypertensive (systolic <140 mm Hg, or diastolic<90 mm Hg) and non-proteinuric (ACR<30mg/mmol) as per DEGREE diagnostic cut-offs [1]. Participants with missing coordinates were excluded from the dataset. Figure 33 shows the final sample size after exclusions.

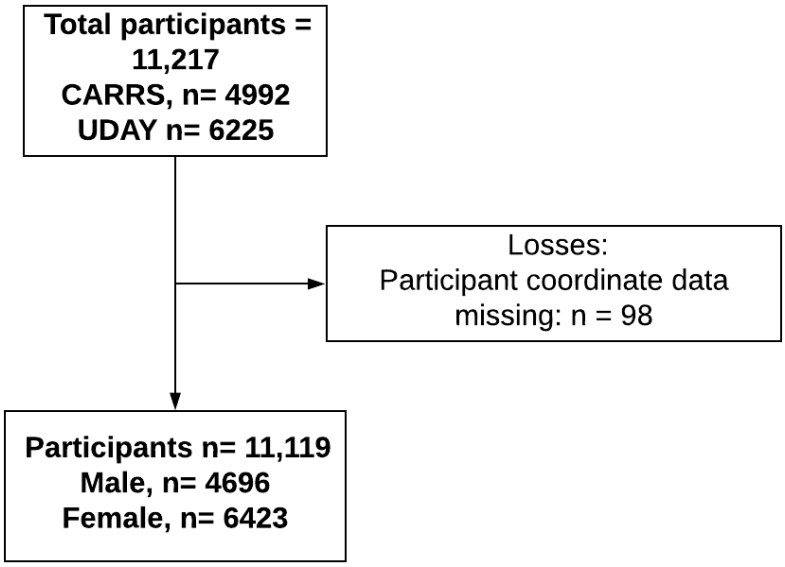


Figure 33. Study flowchart with exclusion criteria for the India population sample

The mean (standard deviation (±SD)) participant age was 41.5(±11.7) years. Mean BMI was 23.6±5.3 kg/m^2, and mean FFM was 41.0±13.0 kg. Mean systolic and diastolic blood pressure was 114.2±11.7 mm Hg, and 73.2±8.7 mm Hg respectively. Mean fasting plasma glucose was 90.2±12.4 mmol/l and the median (interquartile range (IQR) albumin:creatinine ratio (ACR) was 2.6 (1.0-5.0) mg/mmol. Approximately 50% of the participants were employed, and 41% of the population had completed >10 years of formal education.

4.3.2.2 Environmental characteristics

The cross-validation statistics for the heat index model fit (seen in table 28) across the two time periods in India showed an overall good fit, with a marginally better model fit observed for the CARRS time period (R^2= 0.96, RMSE=0.010) than the UDAY time period (R^2= 0.98, RMSE=0.768) (see appendix 10, figures 1 and 2 for model fit regression lines).

Table 28. Empirical Bayesian Kriging cross validation summary for UDAY and CARRS time periods

	India UDAY	India CARRS
Mean average error	0.348	0.501
Root Mean Square Error	0.768	0.010
R^2	0.989	0.969

The heat index across the sampled study zones ranged from 24.4-31.2°C (figure 34-37a) (the bilinear heat index data ranged from 23.6-31.6 °C), and the altitude ranged from 1-391 metres. (figure 34-37c). Approximately 55% of participants lived within 300m of "cropland", (figure 34-37b). For the environmental surfaces over India see figure 38. For the model fit for sensitivity interpolation analyses across India see appendix 11 figures 1-2).

The predicted standard error maps for India follow a similar uncertainty distribution as seen in Sri Lanka, whereby the highest uncertainty of the interpolation model is seen at higher altitudes (figure 28a and 28b for CARRS an UDAY error maps), particularly in the far north and south of the country. In the study zones in question, the uncertainty levels were low. For the CARRS sites in Tamil Nadu and Delhi the error values ranged between 0.193 °C - 0.291 °C and 0.093°C - 0.104 °C respectively. For the UDAY sites in Vishakhapatnam and Sonipat, the error values ranged between and 0.197 °C - 0.297 °C and 0.160-0.249 respectively.

In regions such as the Himalayas, the error was the highest above 2500m (0.89-0.96°C for UDAY time periods) and (0.79-0.94 °C for CARRS time periods) which could be due to the same issue of sparsity of observations in the ERA5 data which (to model most efficiently) required heavier computation of the data which could be being reflected in this heat index surface.

Sensitivity analyses

The comparative results from the validation statistics show that overall bilinear interpolation was the most accurate method in comparison to IDW and kriging (and EBK as shown in table 29). The R^2 was comparable to that of the IDW model (R2 = 0.999), however for bilinear interpolation the RMSE (0.010) an MAE (0.042) had the overall lowest values which would suggest this is a more accurate fit.

Table 29. K fold cross validation Summary of the performance of interpolation methods to predict heat index across India

Interpolation method	Efficiency	Error	
	R^2	RMSE	MAE
IDW	0.999	0.089	0.043
Kriging	0.997	0.014	0.094
Bilinear interpolation	0.999	0.010	0.042
EBK	0.958	0.077	0.385

R^2 = coefficient of determination, RMSE = root mean square error, ME = mean error, MRE = mean relative error, IDW: inverse distance weighting; EBK: Empirical Bayesian Kriging

The Pearson correlation table below (table 30) shows that all interpolation methods are correlated with the original heat index data. Of the simpler methods tested, bilinear interpolation had the highest correlation coefficient (0.73, significant to 0.01). The EBK model with 10-15 neighbours had a marginally higher corelation coefficient with the original data (0.75 significant to 0.01) than that observed for bilinear interpolation. This is likely due to the higher density of data used in this model in comparison to the more simplistic methods.

Table 30. Correlation analysis of interpolation methods across India

	Original HI data	Kriging	Bilinear	IDW	EBK 10-15 neighbours
Pearson correlation	1.00	0.73	0.74	0.73	0.75
p	0.000	0.002	0.001	0.001	0.000

Overall, it could be suggested that, when accounting for error in the validation models and the correlation coefficients, that bilinear interpolation could be a more sensible model to use for the ERA-5 data in India.

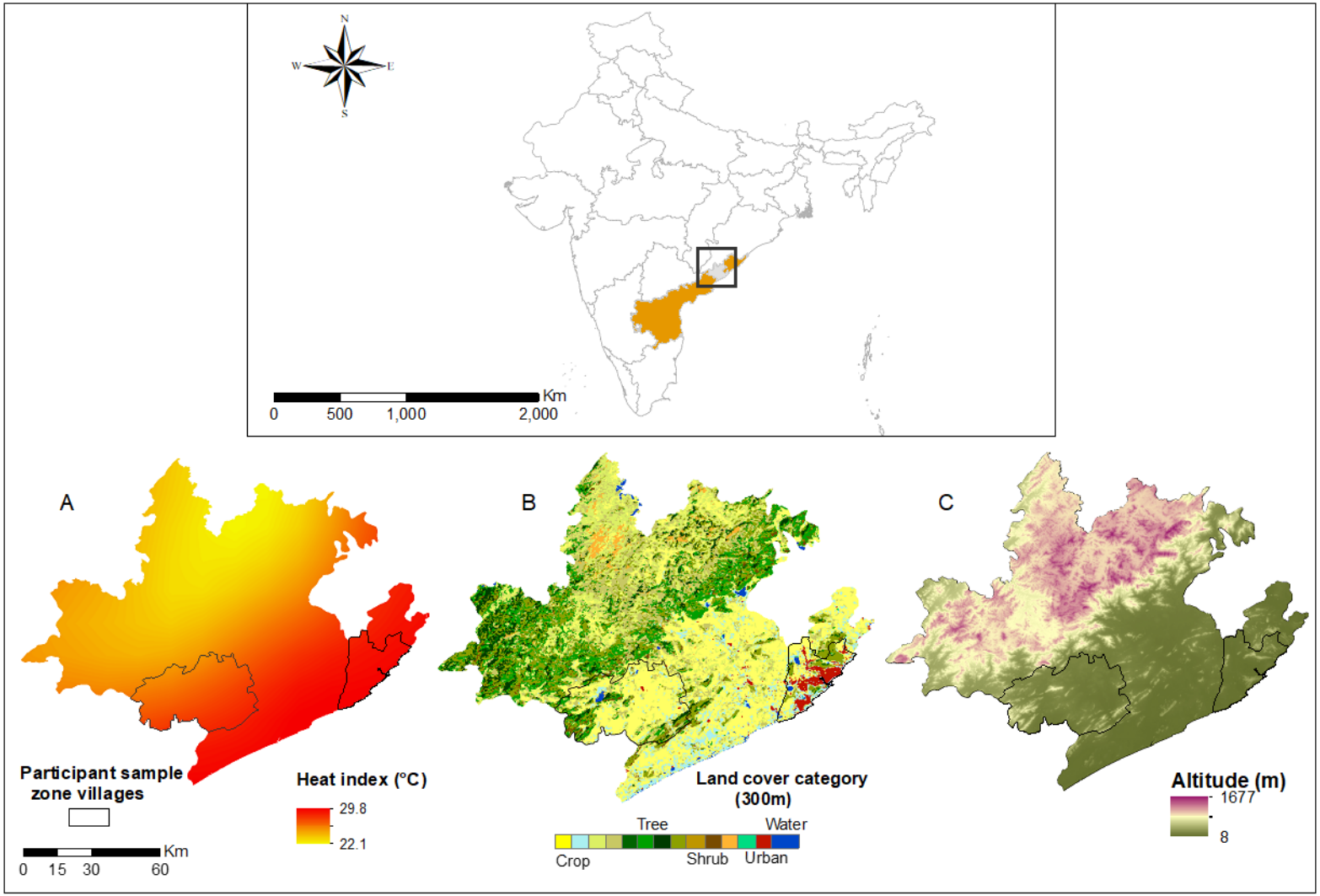


Figure 34. Environmental variable surfaces across Andhra Pradesh study zones showing (a) Heat index, (b) land cover, (c) altitude

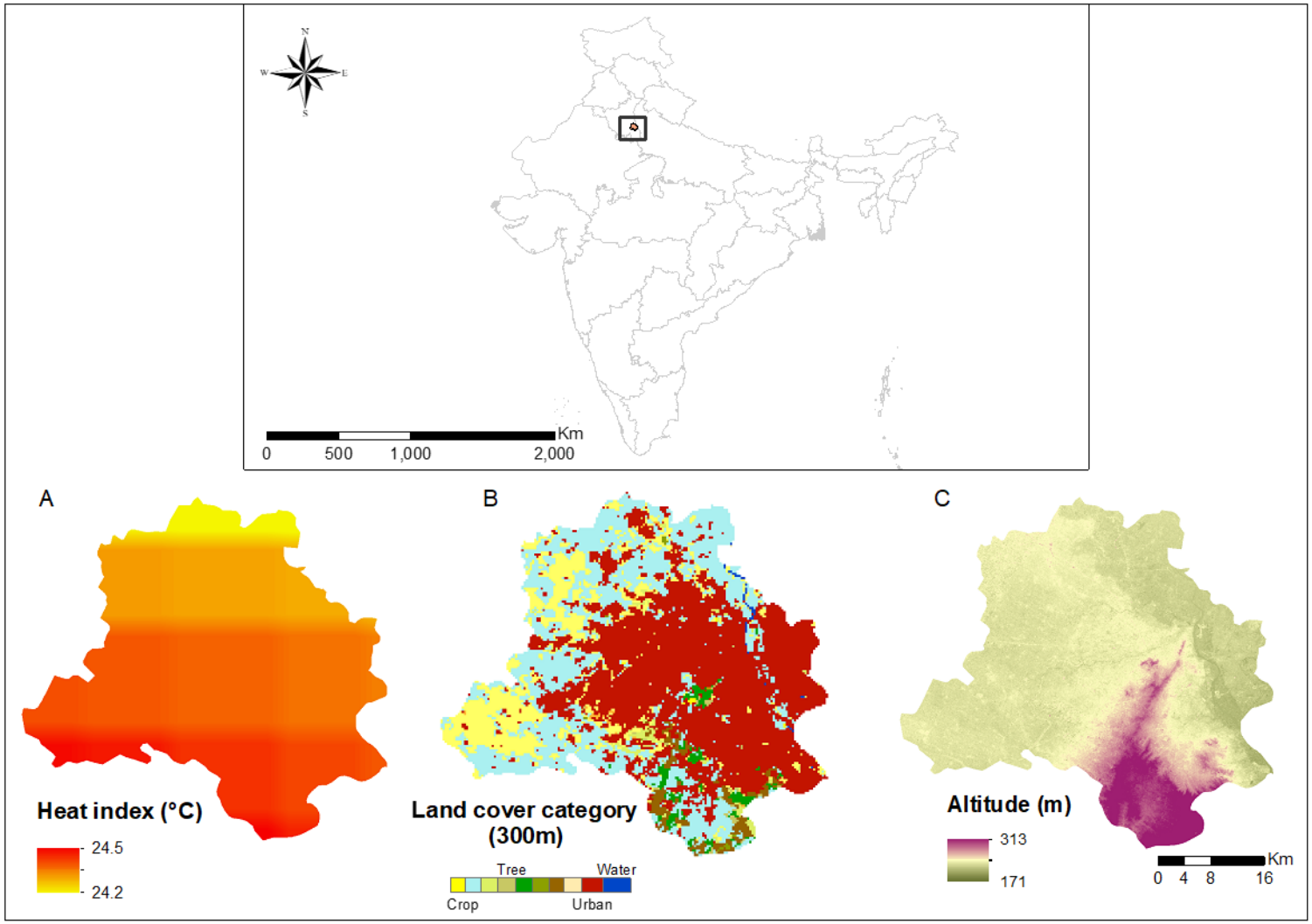


Figure 35. Environmental variable surfaces across Delhi study zone showing (a) Heat index, (b) land cover, (c) altitude

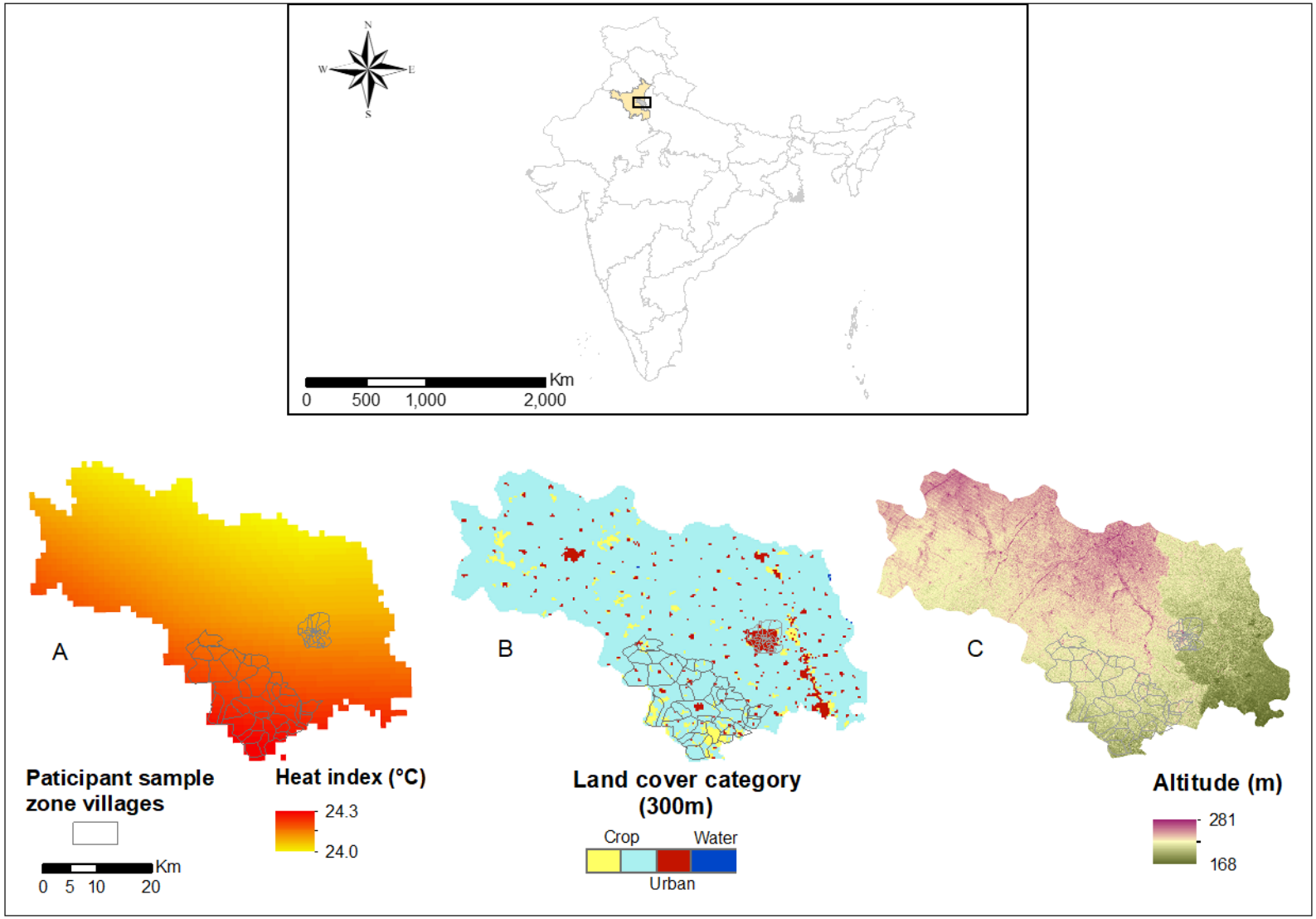


Figure 36. Environmental variable surfaces across Sonipat study zones showing (a) Heat index, (b) land cover, (c) altitude

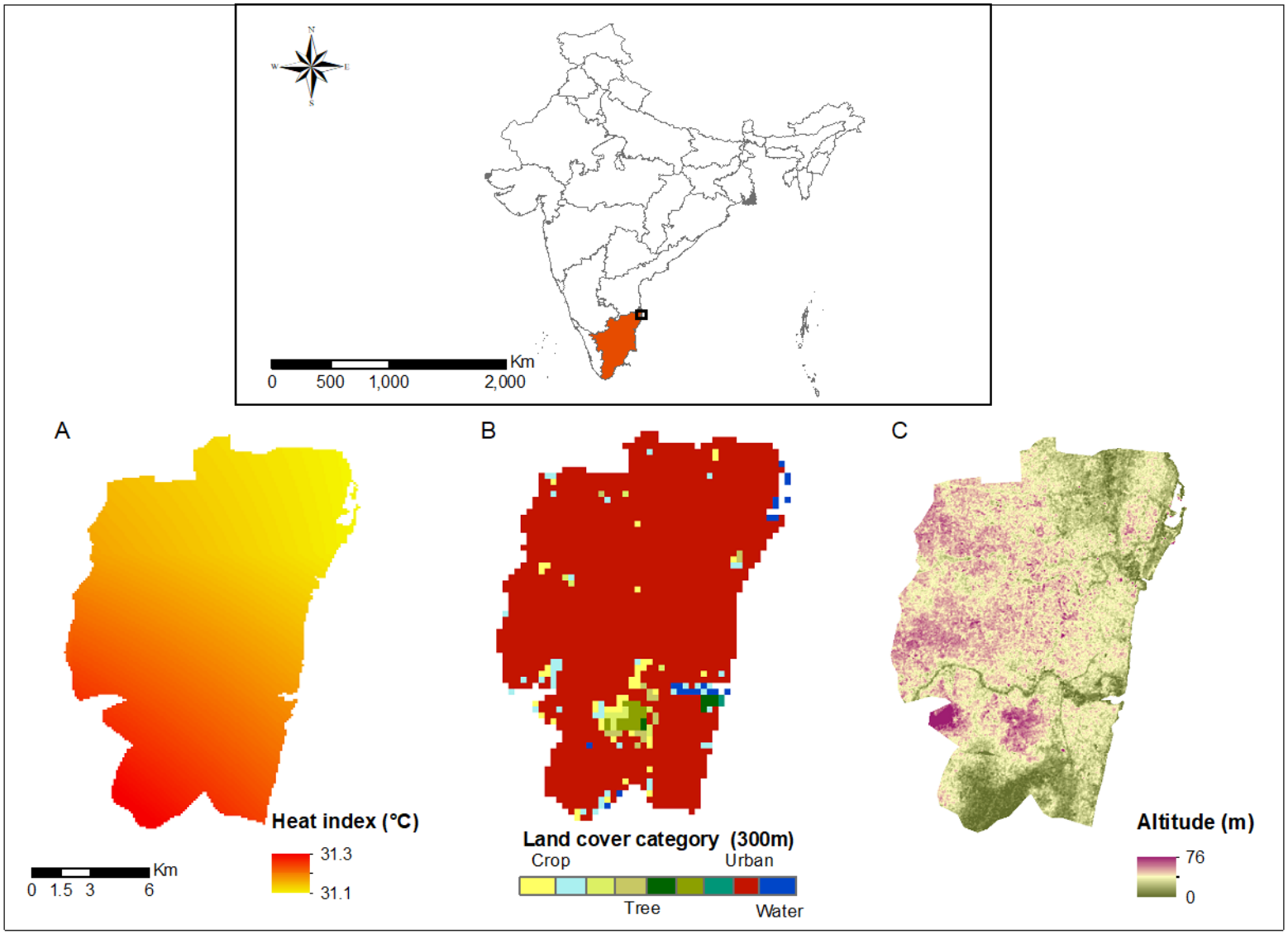


Figure 37. Environmental variable surfaces across Andhra Pradesh study zones showing (a) Heat index, (b) land cover, (c) altitude

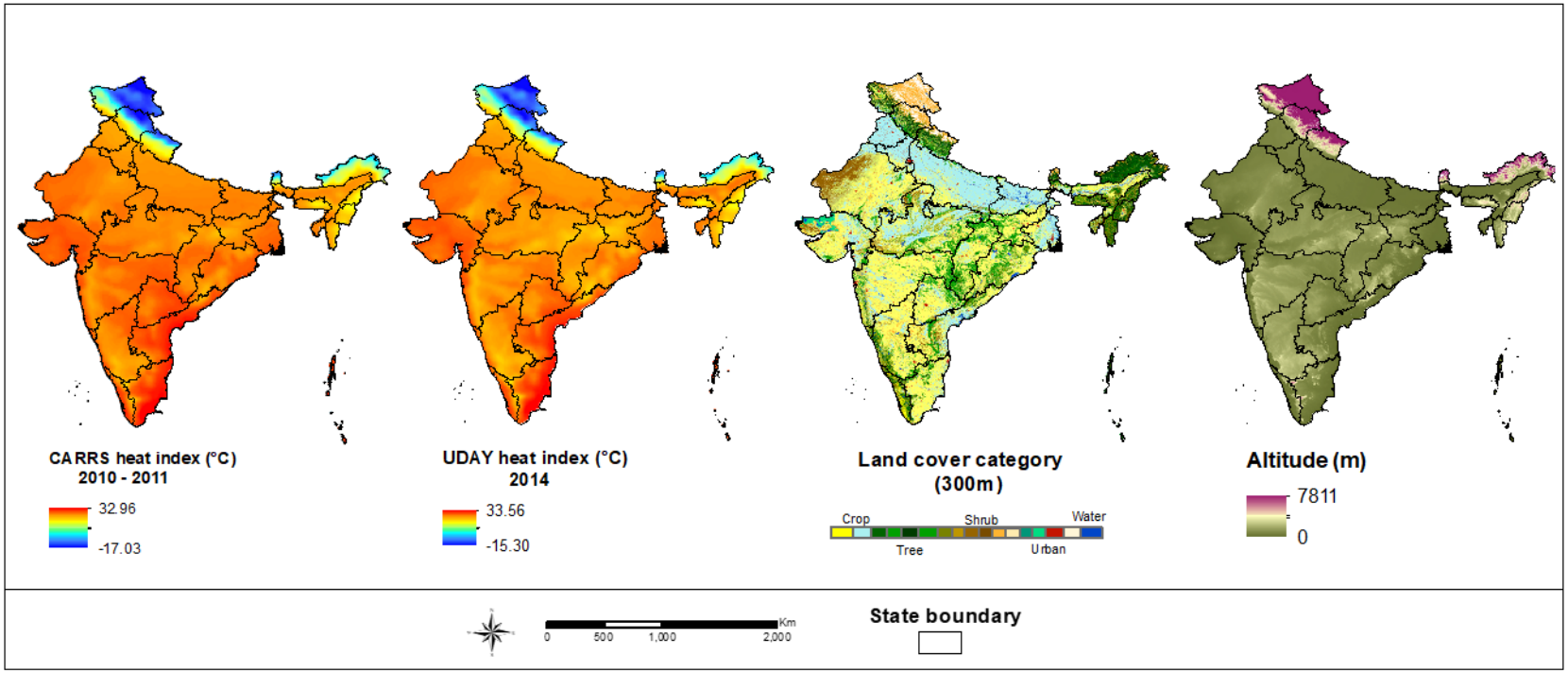


Figure 38. Environmental variable surfaces across India study zones showing (a) Heat index, (b) Altitude, (c) Land cover

4.3.2.3 Mean eGFR and prevalence of low eGFR

The mean eGFR was 105.9(±17.5) mL/min/1.73m^2, and there was an inverse relationship between increasing age, FFM, elevation and income. Males, rural dwellers, those in the vicinity to "cropland" and drinkers and/or smokers had a lower eGFR.

The prevalence of eGFR<60 was 1.4% (95% confidence interval (95%CI) 1.2-1.7). A difference in sex-specific prevalence of eGFR<60 was observed, with 2.0% (95%CI=1.7-2.5) of males versus 1.0% (95%CI=0.8-1.3) of females affected (figure 39). Site-specifically, the highest prevalence of eGFR<60 was observed in Andhra Pradesh (3.2% (95%CI = 2.5-3.8).

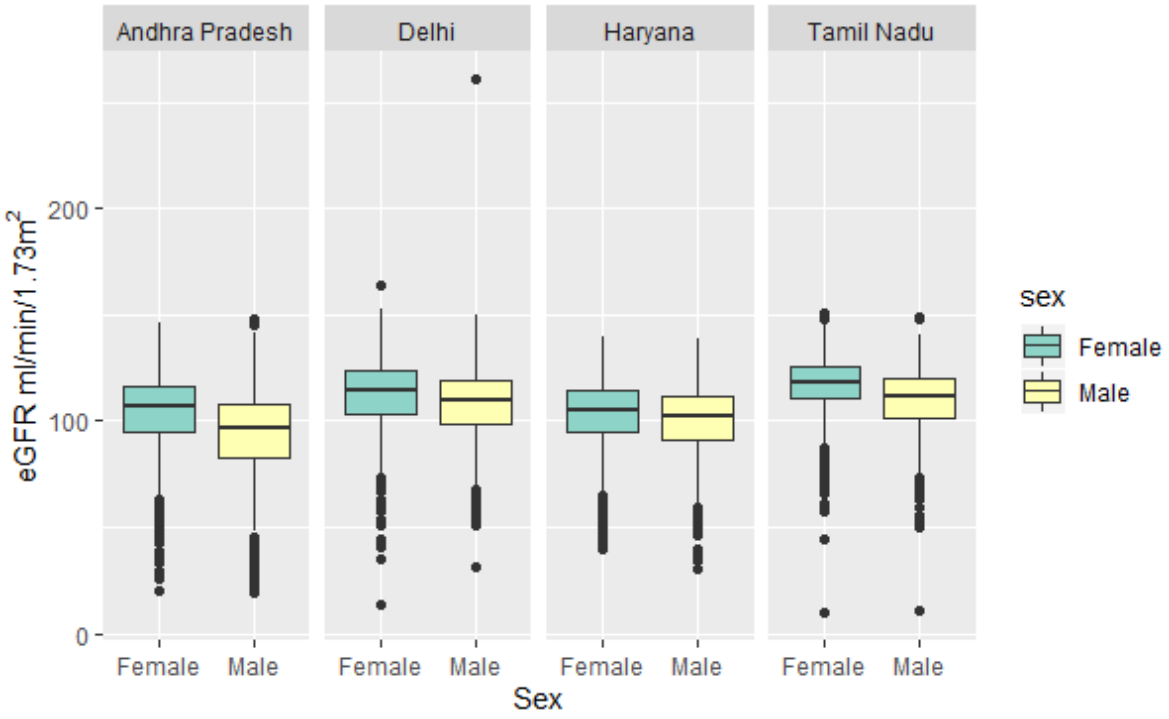


Figure 39. eGFR by sex and state in India

4.3.2.4 Risk factors for reduced eGFR and eGFR<60 using heat index derived from empirical Bayesian kriging

In the initial models there was multicollinearity between altitude (variance inflation factor [VIF] 25.06) and heat index, and north/south latitude with heat index (VIF = 12.06). Like the Sri Lanka analyses, I modelled heat index and altitude separately, and excluded latitude. Heat index can

be used as a proxy for latitude, as broadly temperatures in the north of India are markedly cooler than temperatures in the South. Furthermore, as the study zone urban/rural characteristic variable is very similar to the land cover category ("urban/rural land cover") I excluded the urban/rural characteristic variable to remove duplicate, or very similar variables in the model.

Tables 31 and 32 show linear and logistic regression models which are mutually adjusted for age and sex (minimally adjusted models 1 respectively), incremental modelling with each environmental variable (models 3-5) and models which have been fully adjusted for all socioeconomic and environmental risk factor variables (model 6 respectively). The coefficients do not vary widely between minimally and fully adjusted models I will focus on the results from the fully adjusted models (tables 31 and 32 model 6).

In table 31 model 6, age was a key risk factor for reduced eGFR, with a decrease of 9.18 ml/min/1.73m^2 (95% CI= -9.42, -8.94) per 10-year age increase. Males also had a lower eGFR ([-2.68, (95%CI) = (-3.53, -1.82)]). For the environmental variables, living close to cropland ([-3.39 (95%CI) = (-3.95, -2.83)]) had a negative effect on eGFR. Interestingly, increasing heat index had a weak positive association with eGFR ([0.07 (95%CI= 0.01-0.13). Positive associations were also observed with vegetarianism, education, and lower income.

Like the linear regression model (table 32 model 6), the odds of eGFR<60 increased with age [OR per 10 year age increase (95%CI) = 2.12 (1.83, 2.43)], being male [OR (95%CI) = 1.84 (1.21, 3.25)] and living in the vicinity of cropland [OR(95%CI) = 1.47 (0.12, 1.77)].

Table 31. Associations of sociodemographic, anthropometric, and environmental characteristics with eGFR and eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in India including heat index via Empirical Bayesian kriging, n=11,119

	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]
Variable						
			eGFR			
			Coefficient (95%CI)			
Age [g]	-9.01***	-9.28***	-9.15***	-9.15***	-8.86***	-9.18***
(10-year increase)	(-9.15, -8.72)	(-7.15, -10.62)	(-9.38, -8.91)	(-9.32, -8.91)	(-9.09, -8.63)	(-9.42, -8.94)
Sex [g]						
Male	-3.70***	-2.46***	-2.80***	-2.73**	-3.44***	-2.68***
	(-4.30, -3.31)	(-3.27, -1.64)	(-3.60, -1.99)	(-3.54, -1.92)	(-4.23, -2.64)	(-3.53, -1.82)
Female	REF	REF	REF	REF	REF	REF
Education (years)						
≤5	REF	REF	REF	REF	REF	REF
>5≤10	2.20***	2.11***	2.19*	1.26*	*0.18	1.29***
	(1.56, 2.88)	(2.00, 378)	(1.48, 3.31)	(0.58, 1.93)	(1.81, 0.46)	(0.59, 2.00)
>10	-1.17	-1.09	-1.10	-0.08	-0.13	-0.18
	(-1.51, -0.24)	(-1.99, -0.01)	(-1.59, -0.44)	(-0.59, 0.76)	(-1.81, 0.46)	(-0.52, 0.89)
Occupation						
Employed	REF	REF	REF	REF	REF	REF
Unemployed	1.30***	1.61***	1.57***	1.59**	0.32	1.74***
	(0.91, 2.21)	(1.20, 2.52)	(0.92, 2.23)	(0.94, 2.25)	(-0.33, 0.97)	(1.06, 2.42)
Household monthly income (RS) [i]						
≤30,000	2.96***	2.66***	2.84***	2.70***	1.67***	2.69***
	(2.02, 3.90)	(1.67, 3.68)	(1.86, 3.81)	(1.72, 3.68)	(0.70, 2.64)	(1.68, 3.71)
>30,000	REF	REF	REF	REF		REF
Unknown	-1.27	-1.06	-0.16	-0.27	1.19	-0.31
	(-3.08, 0.53)	(-2.89, 0.76)	(-1.98, 1.65)	(-2.10, 1.54)	(-0.60, 2.99)	(-2.16, -1.54)
BMI (kg/m^2)	-0.60***	-0.61***	-0.67	-0.67	-0.54	-0.82***
5kg/m^2 increase	(-0.84, -0.35)	(-0.83, -0.38)	(-0.90, -0.45)	(-0.89, -0.45)	(-0.76, -0.32)	(-1.20, -0.54)
Fat Free Mass (kg)	-0.34***	-0.16*	-0.12	-0.13	-0.12	-0.13*
5 kg/m2 increase	(-0.43, -0.21)	(-0.27, -0.05)	(-0.23, -0.02)	(-0.24, -0.02)	(-0.23, -0.02)	(-0.24, -0.01)
Healthy lifestyle choices						

Non-drinker and/or smoker	REF	REF	REF	REF	REF	REF
Drinker and/or smoker	0.32	0.39	0.18	0.13	0.32	0.03
	(-0.99, 0.33)	(-0.68, 0.74)	(-0.77, 0.20)	(-0.55, 0.81)	(-0.34, 0.99)	(-0.67, 0.75)
Vegetarian						
Yes	0.25	0.29	1.03***	1.45***	0.29	1.50***
	(-0.29, 0.79)	(0.26, 0.84)	(0.46, 1.59)	(0.82, 2.08)	(-0.43, 1.20)	(-0.85, 2.15)
No	REF	REF	REF	REF	REF	REF
Land cover						
Cropland	-3.40*		-3.40***			-3.39***
	(-3.90, -2.89)		(-3.39, -2.88)			(-3.95, -2.83)
Urban	REF					REF
Altitude						
50m increments	0.12***				0.19	0.15
	(0.07, 0.17)				(0,01, 0.99)	(-0.01, 0.32)
EBK (ºC)						
0.5 increments	0.12***			0.08**		0.07*
	(0.07, 0.17)			(0.02, 0.13)		(0.01, 0.13)

[a] minimal adjustment for age, sex , [b] base variables mutually adjusted; [c] Base variables adjusted for land cover [d] Base variables adjusted for heat index; [e] Base variables adjusted for heat index, [f] Fully adjusted model (base + environmental variables), [g] adjusted for sex, [h] adjusted for age, [i]Exchange rate (RS to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

Table 32. Associations of sociodemographic and anthropometric characteristics with eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in Sri Lanka modelled with Empirical Bayesian Kriging heat data (n=11,119)

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]
	eGFR <60 OR (95%CI)					
Age [g] **(10-year increase)**	2.34*** (1.97, 2.48)	2.19*** (1.92 2.53)	2.15*** (1.89 2.48)	2.19*** (1.92 2.52)	2.16*** (1.89 2.48)	2.12*** (1.83, 2.43)
Sex [h]						
Male	1.43 (1.03, 1.98)	1.73*** (1.14 2.75)	1.79*** (1.14 2.87)	1.73** (1.10 2.75)	1.80** (1.15 2.88)	1.84*** (1.21, 3.25)
Female	Ref	REF	REF	REF	REF	REF
Education (years)						
≤5	REF	REF	REF	REF	REF	REF
>5≤10	0.56** (0.37, 0.85)	0.52*** (0.90 1.21)	0.54*** (0.35 0.82)	0.55* (0.36 0.84)	0.25** (0.35 0.81)	0.58* (0.43, 1.03)
>10	0.28* (0.17, 0.47)	0.22*** (0.24 2.42)	0.25*** (0.14 0.42)	0.25* (0.15 0.43)	0.53 (0.14 0.42)	0.27*** (0.18, 0.58)
Occupation						
Employed	REF	REF	REF	REF	REF	REF
Unemployed	1.17 (0.64, 1.44)	1.04 (0.69 1.57)	1.18 (0.72 1.64)	1.04 (0.70 1.57)	1.10 (0.73 1.67)	1.03 (0.84, 1.95)
Household monthly income (RS) [i]						
≤30,000	1.36 (0.75, 2.79)	0.51 (0.26 1.13)	0.51 (0.26 1.11)	0.51 (0.26 1.12)	0.50 (0.26 1.10)	0.55 (0.28, 1.26)
>30,000	REF	REF	REF	REF		REF
Unknown	1.62 (0.54, 4.52)	0.48 (0.145 1.43)	0.43 (0.14 1.29)	0.47 (0.15 1.42)	0.43 (0.73 1.67)	0.46 (0.14, 1.43)

BMI (kg/m^2)	0.80*	0.88	0.88	0.88	0.89	0.84
5kg/m^2 increase	(0.69, 0.94)	0.78 1.04)	0.77 1.05)	(0.78 1.04)	(0.77 1.05)	(0.71, 1.08)
Fat Free Mass (kg)	0.91	1.01	1.00	1.01	1.00	0.99
5 kg/m2 increase	(1.03, 1.23)	(0.63 1.35)	(0.93 1.05)	(0.95 1.04)	(0.93 1.05)	(0.88, 1.05)
Healthy lifestyle choices						
Non-drinker and/or smoker	REF	REF	REF	REF	REF	REF
Drinker and/or smoker	0.96	0.91	0.85	0.93	0.86	0.84
	(0.70, 1.50)	(0.63 1.35)	(0.58 1.27)	(0.64 1.38)	(0.59 1.28)	(0.58, 1.28)
Vegetarian						
Yes	0.48***	0.61*	0.57***	0.65***	0.63	0.58*
	(0.32, 0.66)	(0.42 0.89)	(0.39 0.84)	(0.43 0.99)	(0.40 1.00)	(0.57, 1.62)
No	REF	REF	REF	REF	REF	REF
Land cover						
Cropland	1.90***		1.55*			1.47*
	(1.36, 2.70)		(1.09, 2.27)			(0.12 1.77)
Urban	REF					REF
Altitude						
50m increments	0.90*				0.98	0.93
	(0.83, 0.99)				(0.85 1.08)	(0.83, 1.06)
EBK heat index (°C)						
0.5 increments	0.84***			1.07		0.98
	(0.78, 0.91)			(0.90 1.30)		(0.79, 0.93)

[a] minimal adjustment for age, sex, [b] base variables mutually adjusted; [c] Base variables adjusted for land cover [d] Base variables adjusted for heat index; [e] Base variables adjusted for heat index, [f] Fully adjusted model (base + environmental variables), [g] adjusted for sex, [h] adjusted for age, [i]Exchange rate (RS to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/

4.3.2.5 Risk factors for reduced eGFR and eGFR<60 using heat index derived from bilinear interpolation

In the linear regression model using bilinear interpolated heat index data, results remained largely unchanged from the model using the original EBK derived heat index data, with negative associations remaining with increasing age, being male, and living in vicinity to cropland (table 33 model 6).

The results in the logistic regression model (table 34 models 1-6) were also largely unchanged from the results seen using the EBK derived heat index data, with similar increased risks observed with age, sex and vicinity to cropland.

Table 33. Associations of sociodemographic, anthropometric, and environmental characteristics with eGFR in participants without diabetes, hypertension and heavy proteinuria in India including heat index via bilinear interpolation, n=11,119

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]
	eGFR Coefficient (95%CI)					
Age [g]	-9.01***	-9.16***	-9.15***	-9.28***	-9.25***	-9.24***
(10-year increase)	(-9.15, -8.72)	(-9.51, -9.05)	(-9.38, -8,91)	(-9.51, -9.05)	(-9.48, -9.02)	(-9.36, -8.89)
Sex [h]						
Male	-3.70***	-2.46***	-2.80***	-2.44***	-2.45***	-2.41***
	(-4.30, -3.31)	(-3.27, -1.64)	(-3.06, -1.99)	(-3.25, -1.62)	(-3.26, -1.64)	(-3.58, -1.94)
Female	REF	REF	REF	REF	REF	REF
Education (years)						
≤5	REF	REF	REF	REF	REF	REF
>5≤10	2.20***	2.32**	1.37**	2.10**	2.00**	2.00***
	(1.56, 2.88)	(1.45, 2.79)	(0.69, 2.04)	(1.43, 2.77)	(1.32, 2.67)	(0.69, 2.04)
>10	-1.17	-1.32	-018	1.01	1.01	1.00
	(-1.51, -0.24)	(0.43, 2.79)	(-0.49, 0.86)	(0.40, 1.74)	(0.34, 1.68)	(-0.49, 0.87)
Occupation						
Employed	REF	REF	REF	REF	REF	REF
Unemployed	1.30***	1.86***	1.57***	1.88***	1.82***	1.82***
	(0.91, 2.21)	(1.20, 2.52)	(0.92, 2.23)	(1.22, 2.54)	(1.16, 2.48)	(0.90, 2.23)
Household monthly income (RS) [i]						
≤30,000	2.96***	2.66***	2.84**	2.63***	2.52***	2.47***
	(2.02, 3.90)	(1.67, 3.64)	(1.86, 3.81)	(1.65, 3.62)	(1.53, 3.51)	(1.81, 3.77)
>30,000	REF					REF
Unknown	-1.27	-1.06	-0.16	-1.09	-1.09	-0.11
	(-3.08, 0.53)	(-2.89, 0.76)	(-1.98, -1.65)	(-2.92, 0.73)	(-2.92, 0.73)	(-1.71, -0.02)
BMI (kg/m^2)	-0.60***	-0.61**	-0.67**	-0.61**	-0.60**	-0.61***
5kg/m^2 increase	(-0.84, -0.35)	(-0.83, -0.05)	(-0.90, -0.45)	(-0.83, -0.38)	(-0.83, -0.38)	(-0.90, -0.45)

Fat Free Mass (kg)	-0.34***	-0.16*	-0.12*	-0.16*	-0.16*	-0.16*
5 kg/m2 increase	(-0.43, -0.21)	(-0.27, -0.05)	(-0.23, -0.02)	(-0.27, -0.05)	(-0.27, -0.06)	(-0.23, -0.02)
Healthy lifestyle choices						
Non-drinker and/or smoker	REF	REF	REF	REF	REF	REF
Drinker and/or smoker	0.32	0.32	0.07	-0.30	-0.27	0.25
	(-0.99, 0.33)	(-1.00 0.36)	(-0.60, 0.75)	(-0.99, 038)	(-0.96, 0.40)	(-0.58, 0.79)
Vegetarian						
Yes	0.25	0.29	1.03*	0.43	0.62*	0.60*
	(-0.29, 0.79)	(-0.26, 0.84)	(0.46, 1.59)	(-0.24, 1.11)	(0.03, 1.22)	(0.47, 1.68)
No	REF		REF			REF
Land cover						
Cropland	-3.40*		-3.40***			-3.42***
	(-3.90, -2.89)		(-3.93, -2.88)			(-3.97, -2.88)
Urban	REF					REF
Altitude						
50m increments	0.12***			0.05		0.15
	(0.07, 0.17)			(-0.21, 0.09)		(-0.01, 0.32)
Bilinear heat index (ºC)						
0.5 increments	0.08***				0.06***	0.02*
	(0.04, 0.13)				(0.02, 0.11)	(0.01, 0.06)

[a] minimal adjustment for age, sex, [b] base variables mutually adjusted; [c] Base variables adjusted for land cover [d] Base variables adjusted for heat index; [e] Base variables adjusted for heat index, [f] Fully adjusted model (base + environmental variables), [g] adjusted for sex, [h] adjusted for age, [i]Exchange rate (RS to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmo

Table 34 Associations of sociodemographic and anthropometric characteristics with eGFR<60 in participants without diabetes, hypertension and heavy proteinuria in Sri Lanka modelled with bilinear interpolation heat data (n=11,119)

	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]
Variable	eGFR <60 OR (95%CI)					
Age [g] (10-year increase)	2.34*** (1.97, 2.48)	2.19*** (1.92 2.52)	2.15*** (1.89 2.48)	2.19*** (1.92 2.52)	2.19*** (1.92 2.52)	2.16*** (1.89, 2.50)
Sex [h]						
Male	1.43 (1.03, 1.98)	1.66* (1.10 2.75)	1.79** (1.14 2.87)	1.73*** (1.11 2.75)	1.73*** (1.10 2.75)	1.83*** (1.15, 2.97)
Female	Ref	REF	REF	REF	REF	REF
Education (years)						
≤5	REF	REF	REF	REF	REF	REF
>5≤10	0.56** (0.37, 0.85)	0.49* (0.32 0.73)	0.54* (0.35 0.82)	0.48* (0.32 0.73)	0.48* (0.32 0.73)	0.25* (0.34, 0.81)
>10	0.28* (0.17, 0.47)	0.22* (0.13 0.37)	0.25* (0.14 0.42)	0.21* (0.13 0.36)	0.21* (0.13 0.37)	0.33*** (0.26, 1.12)
Occupation						
Employed	REF	REF	REF	REF	REF	REF
Unemployed	1.17 (0.64, 1.44)	1.04 (0.69 1.57)	1.08 (0.72 1.64)	1.04 (0.70 1.58)	1.04 (0.69 1.57)	1.12 (0.74, 1.71)
Household monthly income (RS) [i]						
≤30,000	1.36 (0.75, 2.79)	0.50 (0.26 1.13)	0.51 (0.26 1.11)	0.51 (0.26 1.12)	0.52 (0.26 1.13)	0.51 (0.26, 1.12)
>30,000	REF	REF	REF			REF
Unknown	1.62 (0.54, 4.52)	0.47 (0.15 1.43)	0.43 (0.14 1.29)	0.47 (0.15 1.41)	0.48 (0.15 1.43)	0.39 (0.14, 1.43)

BMI (kg/m^2)	0.80*	0.89*	0.88	0.89	0.88	0.89
5kg/m^2 increase	(0.69, 0.94)	(0.78 1.04)	(0.77 1.05)	(0.78 1.04)	(0.78 1.04)	(0.78, 1.07)
Fat Free Mass (kg)	0.91	1.01	1.00	1.01	1.01	1.00
5 kg/m2 increase	(1.03, 1.23)	(0.94 1.05)	(0.93 1.05)	(0.94 1.05)	(0.94 1.05)	(0.92, 1.05)
Healthy lifestyle choices						
Non-drinker and/or smoker	REF	REF	REF	REF	REF	REF
Drinker and/or smoker	0.96	0.91	0.85	0.92	0.91	0.88
	(0.70, 1.50)	(0.63 1.35)	(0.58 1.27)	(0.63 1.36)	(0.62 1.35)	(0.59, 1.32)
Vegetarian						
Yes	0.48***	0.61*	0.57*	0.64*	0.61*	0.59*
	(0.32, 0.66)	(0.42 0.89)	(0.39 0.84)	(0.41 1.01)	(0.41 0.91)	(0.39, 0.89)
No	REF	REF	REF			REF
Land cover						
Cropland	1.90***		1.55*			1.59*
	(1.36, 2.70)		(1.09 2.27)			(1,09, 2.36)
Urban	REF					REF
Altitude						
50m increments	0.90*			0.98		0.60
	(0.83, 0.99)			(0.88 1.09)		(0.11, 2.36)
HI bilinear interpolation (ºC)						
0.5 increments	0.84***				0.99	1.00
	(0.78, 0.91)				(0.97 1.03)	(0.97, 2.36)

[a] minimal adjustment for age, sex, [b] base variables mutually adjusted; [c] Base variables adjusted for land cover [d] Base variables adjusted for heat index; [e] Base variables adjusted for heat index, [f] Fully adjusted model (base + environmental variables), [g] adjusted for sex, [h] adjusted for age, [i]Exchange rate (RS to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/

4.3.2.6 Multiple regression models

Variables associated with low eGFR and eGFR<60 in preliminary analyses were added into the multiple regression model, in addition to all environmental variables (table 35). Age, sex, and vicinity to cropland remained key risk factors for low eGFR and risk of eGFR<60 (models 1-2).

Table 35. Multiple regression models of sociodemographic and environmental characteristics associated with eGFR and eGFR<60 India

Variable	Model 2 Fully adjusted eGFR Coefficient (95%CI)[b];	Model 3 Minimal adjustment eGFR <60 OR (95%CI)[a];
Age **(10-year increase)**	-9.13*** (-9.37, -8.90)	2.14*** (1.90, 2.44)
Sex		
Male	-4.00*** (-4.54, -3.46)	1.61** (0.14, 2.29)
Female	REF	Ref
Education (years)		
≤5	REF	REF
>5≤10	1.46*** (0.77, 2.46)	0.59* (0.39, 0.91)
>10	0.03 (-0.67, 0.68)	0.31* (0.18, 0.52)
Heat index EBK (ºC) **0.5 increment**	0.08 (0.02, 0.14)	0.98 (0.94, 1.03)
Heat index bilinear interpolation (ºC) **0.5 increment**	0.03 (0.01, 0.07)	1.01 (0.97, 1.04)
Land cover		
Cropland	-3.50*** (-4.06, -2.95)	1.43* (0.98, 2.11)
Urban	REF	
BMI **5kg/m^2 increase**	0.86 (-1.24, -0.74)	0.85* (0.73, 1.01)
Vegetarian		REF
Yes	1.60*** (0.96, 2.24)	0.56*** (0.37, 0.85)
No	REF	REF
Altitude **50m increment**	0.26*** (0.13, 0.40)	0.88* (0.94, 1.03)

4.3.2.7 Linear mixed models

The caterpillar plot below (figure 40) shows the results of the linear mixed model for India which has been modelled with a hierarchy of state:urban/rural area as the random structures, and environmental variables heat index, altitude, and crop land as the fixed effects. The results indicate that both rural and urban Andhra Pradesh (southern India) have the lowest eGFR across the Indian study zones. In Haryana (northern India) the rural areas have a low eGFR, whereas the urban area eGFR does not deviate from the overall mean. Delhi (northern India) and urban Tamil Nadu (southern India) had the highest mean eGFR.

The fixed effects table (table 36) shows the effects of the environmental variables, altitude, and crop land and heat index (including both the EBK method and bilinear interpolation derived heat index data which has been appended at the end of the model) on eGFR across all study zones. Altitude had a significant negative impact on eGFR (-1.12 (-2.16, -0.07)), as did land cover category (-1.29 (-0.34, -2.37)). There was no association observed with heat index and eGFR for either of the heat index datasets

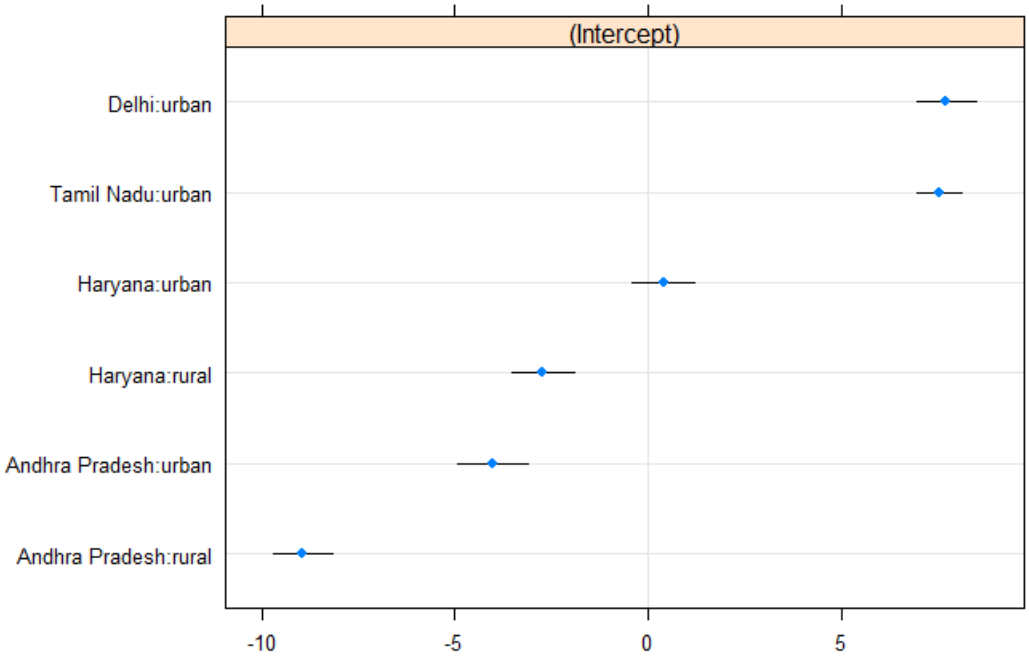


Figure 40. Linear mixed model caterpillar plot of eGFR in study zones in India accounting for land cover, heat index and altitude. Intercept denotes the overall mean eGFR across the study zones; blue circles represent the average eGFR for each zone; black bars represent 95% confidence intervals.

Table 36. Linear mixed model for altitude, heat index and land cover in India

Variable	Regression coefficient	Confidence interval (95%)	P value
Heat index EBK (°C)	0.19	-0.59, 0.47	0.40
Land cover (crop)	-1.29	-2.37, -0.34	<0.001***
Altitude (m)	-1.12	-2.16, -0.07	0.01**
Heat index bilinear interpolation (°C)	0.20	-0.41, 1.26	0.40

4.3.2.8 Linear mixed model International comparison

Like the regression analyses above, the altitude and heat index variables were modelled separately in order to determine independent associations with eGFR. The international comparison between Sri Lanka and India indicates that study zone 3 in Sri Lanka has the lowest eGFR (figure 41), followed by zones 1 and 2. In India, Andhra Pradesh has the lowest eGFR rank, and Delhi has the highest, followed by Haryana and Tamil Nadu. These results clearly show a Sri Lanka < India gradient in eGFR, however it must be noted that the sample sizes between the countries differ largely (Sri Lanka n= 3296; India n= 6423). This difference could have affected the results depicted in figure 41 due to potential misclassification of low eGFR as the Indian data was not DEGREE standardized. It is therefore challenging to conduct any further sensitivity analyses on these data to investigate this pattern. These analyses can be conducted and re-tested once the DEGREE data from India is available in order to glean further information on the effect of samples size on the LMM results.

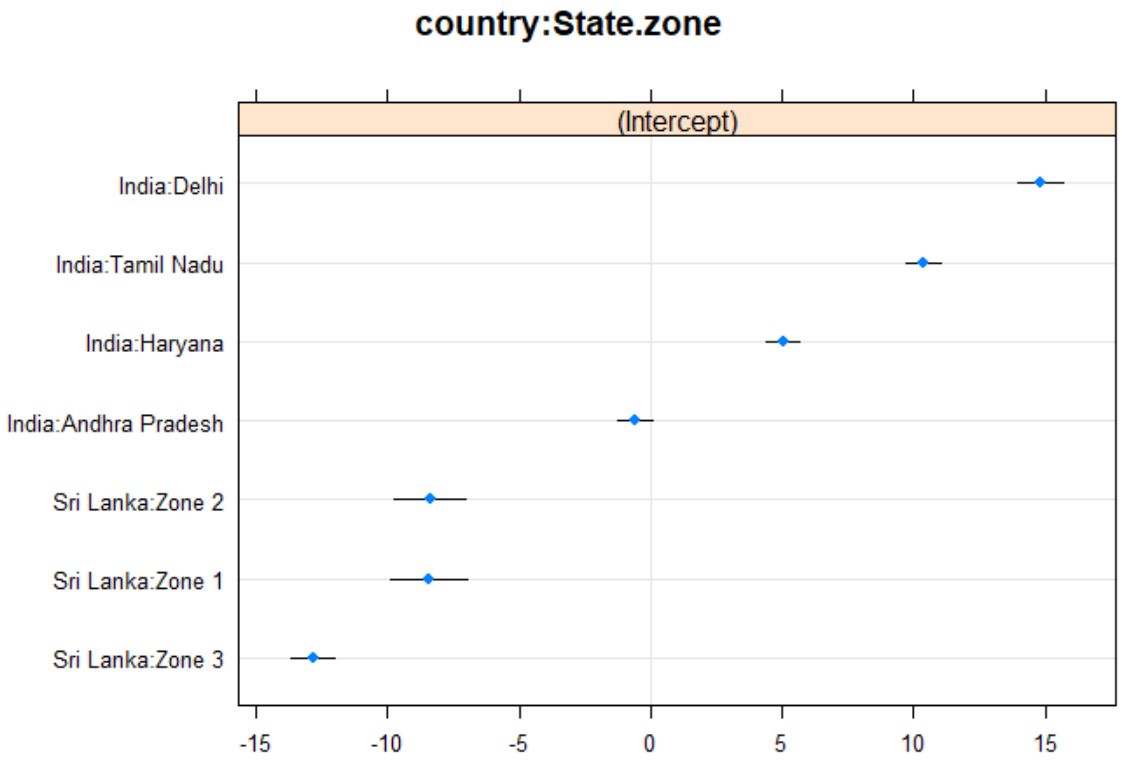


Figure 41. Linear mixed model caterpillar plot of eGFR between study zones 1-3 in Sri Lanka and study zone states in India. Intercept denotes the overall mean eGFR across the study zones; blue circles represent the average eGFR for each zone; black bars represent 95% confidence intervals.

Across both countries and regions, heat index has a minimal positive effect on eGFR, and no association was observed for land cover category. Table 37 shows the association of altitude and land cover and heat index (bilinear heat index appended) on eGFR across India and Sri Lanka study zones. There was a significant negative effect observed with increasing altitude (-3.10 (-3.98, -2.22), and no association was observed between eGFR and land cover.

Table 37. Fixed effects table of environmental factor (heat index, Land cover and altitude) effect on eGFR across Indian and Sri Lankan study zones

Variable	Regression coefficient	Confidence interval (95%)	P value
Heat index EBK (°C)	0.62	0.35, 0.88	<0.001***
Land cover (crop)	-0.58	-1.46, 0.29	0.19
Altitude (m)	-3.10	-3.98, -2.22	<0.001***
Heat index bilinear (°C)	0.46	0.39, 0.53	<0.001***

4.4 Discussion

The results from this exploratory geographical analysis show the prevalence of CKDu is markedly higher in Sri Lankan populations (5.8%) than Indian (1.4%), and that environmental risk factors of low eGFR and eGFR<60 vary by country.

The Sri Lanka study zones are situated in low-lying areas with most participants living at altitudes >100m. Heat index temperatures ranged from 32.3-33.3°C, and the entire population lived in the vicinity of crop and tree cover categories. In India, the altitude ranged from sea level to 391m, and over half of the study population lived at altitudes <100m. Furthermore, as the study zones span the entire country, the heat index temperatures had a wider range of 24.1-31.2°C. As expected, the land cover categories were a mix of urban and cropland, and approximately 55% of the population lived in 300m vicinity of crops. In general, these conditions (higher temperatures, lower altitudes and a majority living in and around agricultural land) concur with the environmental characteristics associated with CKDu across the literature [10, 17, 256, 257].

As the environmental characteristics of the Sri Lankan and Indian study zones are relatively similar, the wide disparity in prevalence of CKDu in neighbouring countries could be due to there simply being a more widespread issue in Sri Lanka, which could be due to factors such as the environmental exposures explored in this chapter, in addition to other factors such as genetic predisposition in this population sample. This could also however be due to the fact that the data sources for each country differ widely – the Sri Lanka dataset is DEGREE standardized and therefore has different, more accurate measures in place for renal function

estimation which may be more robust that those used for the Indian data. This factor will require further exploration upon the completion of the DEGREE study currently being conducted in India.

Empirical Bayesian Kriging produced a good fit for heat index surfaces over Sri Lanka and India, all of which had an R^2 value >0.90. The good fit was likely due to the high density of observation points across each country. This method proved sufficient at producing predictive surfaces for heat index using ERA-5 data, however sensitivity analyses showed that use of a more simplistic model (notably bilinear interpolation) was likely the most appropriate model to create predictive surfaces. The EBK model was made using the entire EERA-5 modelled data across India and Sri Lanka which was then scaled down to the study zone level. Whilst the R^2 value was very high for these models, the error was higher than those observed across the simpler interpolation models which would suggest that in this case these simpler methods would be more appropriate to use. The sensitivity analyses were conducted by lowering the nearest neighbour search radius from 10-15 neighbours to 2-4, and I also restricted the heat index data coverage to within 100m of the study zones across each country in order to create a more reliable model. Overall, across both India and Sri Lanka, bilinear interpolation created the most accurate predictive surfaces, and this more simplistic method is the most appropriate to use in this instance.

In Sri Lanka variables associated with both low eGFR and risk of eGFR<60 was age and drinking and/or smoking and conducting chena cultivation. In linear regression models and mixed models, the environmental variables associated with low eGFR were increasing heat index, altitude, farming vegetables and chena cultivation.

In linear regression models, the negative association between both heat index and altitude is unexpected, as they generally have an inverse relationship, as reported in the eGFR-altitude studies in Central America [14]. I conducted further investigation on these findings and observed no multicollinearity between the variables when modelled together, therefore this observation could be due to the following reasons. Study zones 1-3 span a relatively small area approximately 50km, across which there is a small South-North (or zone 1-3) heat index gradient (1^{o}C increase). It may be that even across this small area the latitude effect is stronger than any altitude effect and could account for the heat index -eGFR negative association. It

must be noted that there were large changes in effect estimates for agricultural workers after full adjustment of variables in linear regression models (tables 21-24, models 1-6).

In the case of agricultural labourers, the introduction of the chena cultivation variable caused the change observed in the "agricultural" occupation category. Although there was no multicollinearity between these variables, this could again be due to the very high proportion of agricultural workers who reported practicing chena cultivation (74%) which have an eGFR <60, and furthermore there was a slightly raised variance inflation factor between agricultural labourers and chena cultivation practices (4.01 and 4.16 respectively) which could have caused the association between agricultural labourers and reduced eGFR to be lost. I modelled the fully adjusted variables again in the absence of chena cultivation and the association between agricultural labour and reduced eGFR was restored, which would suggest that the presence of both these variables in one model are causing spurious results.

I did not adjust for study zone in the main models as the aim was to model the risk factors for the study area as a whole; however, it was important to explore the zonal risk factors in a separate model to ascertain if there are any differences in site-specific risk factors. Across zones 1 and 2, age was the only risk factor associated with eGFR<60 which might suggest that the low eGFR we observe in these zones could be normal age-related renal decline. In zone 3 age, drinking and/or smoking and drinking water from tube wells and chena cultivation were associated with eGFR<60. In Sri Lanka drinking water is a widely debated hypothesis as a causative factor for CKDu, with studies purporting that regions with the highest CKDu prevalence use wells as their primary water source[21, 129, 258]. Although hydro chemical characterization of groundwater analyses in Sri Lanka have not provided conclusive evidence on the causative nature of water sources[130, 213, 215], this hypothesis still remains largely supported in the Sri Lanka research field, and this finding could further support this debate.

In India, age and sex were associated with both low eGFR and increased risk of eGFR<60. Of the environmental risk factors, vicinity to cropland was associated with both low eGFR and risk of eGFR<60. This finding corroborates with a previous study in El Salvador whereby proximity to specific crop types was significantly associated with suspected CKDu mortality and hospital admissions rates[10]. Although this dataset does not contain this high-resolution crop detail, it could help geographically highlight specific sub-communities which could be at a higher risk and help direct future observational and treatment plans in affected regions.

In multiple regression models for India age, sex and proximity to cropland remained associated with low eGFR and risk of eGFR<60, whilst in Sri Lanka key risk factors were age, drinking/smoking and chena cultivation.

The prevalence of CKDu is reportedly highest among communities that engage in chena cultivation and paddy farming, which is most often practiced in the dry zone including the North Central Provinces (NCP) of Sri Lanka[23, 73, 164], which is where the DEGREE study was based. This chena risk factor finding corroborates with this observation, and also a previous study [23], which found chena cultivators at a higher risk of CKDu than other types of cultivation practices in the region. Chena cultivation is considered as a strenuous farming practice as it involves the "slash and burn" of jungle or forest before cropping which could cause renal injury by inducing hyper-osmolarity and damage to the proximal tubules via the fructokinase pathway (as described in chapter 1) [81, 259]. In Central America similar "slash and burn" cultivation of sugar cane farming is considered a major risk factor for CKDu[13, 29, 82, 186], thus this finding of a similarly strenuous type of labour suggest this could be a common risk factor for CKDu globally.

I modelled the data using both EBK and bilinear interpolation-generated heat index data to compare how differing methods (and resulting errors) could affect linear and logistic regression results. Overall, there were no marked changes observed when adjusting for each heat index data model across Sri Lanka and India, however, the selection of the most accurate model is paramount to reducing the risk of exposure misclassification across a population.
The LMM results were interesting as they presented very different observations from the linear and logistic regression models, as they account for environment and study site only. In Sri Lanka, chena cultivation and rice farming, heat index, altitude and use of pesticides had the strongest negative effect on eGFR. Drinking from water sources other than groundwater had a positive effect on eGFR. There is no information on what these other water sources are, however, there is an increasing amount of research investigating the remediating effects of reverse osmosis and bottled water in CKDu endemic regions[260, 261]. In India, both the urban and rural sites in Andhra Pradesh had the lowest ranking eGFR, followed by rural Haryana. Urban Delhi and Tamil Nadu had the highest ranking eGFR. Vicinity to cropland and increasing altitude both had a negative association with eGFR; however, no effect was observed with heat index.

In the linear mixed model comparison of heat index, altitude and land cover category in India and Sri Lanka, Sri Lanka had the lowest ranking eGFR, with zones 1-3 all below ranking below the mean eGFR. Zone 3 had the lowest ranking eGFR, whilst in zones 1 and 2 there was no mean difference observed. In India, Andhra Pradesh had the lowest ranking eGFR, however the confidence intervals cross the intercept, thus the mean does not deviate from the overall average. The remaining Indian states all ranked higher than the average eGFR, with Delhi ranking the highest. Interestingly, in this model increasing heat index had a positive association with eGFR, which could be due to the fact that Sri Lanka has a lower maximum heat index than India, and there is generally a higher prevalence of CKDu in Sri Lanka, thus a spurious (albeit weak) protective effect could be observed in this instance. Results including bilinear interpolation data showed no marked difference in environmental coefficients across any LMMs.

This study has some limitations including the non-comparability of the biomarker data in the Sri Lanka and India datasets. This however might not pose significant issues in the geographical study setting, as at present there is no consensus as to whether the disease is histologically the same across reporting regions; thus, using diverse data sources to investigate broad patterns of environmental risk factors is valuable at this exploratory stage.

Finally, the study zones in Sri Lanka were entirely rural and selected based on a priori hypotheses of high CKDu prevalence. To avoid the “streetlight effect” and reduce selection bias, it would be optimal to compare this rural region to an urban and /or mixed land region to understand if there is an urban/rural difference in eGFR in Sri Lanka.

4.4.1 Conclusions

In conclusion, the results from this ecological analysis suggest that environmental risk factors which have been hypothesised as universal to all CKDu regions, are instead geographically independent and challenge environmental hypotheses which are widely debated in the literature.

Sri Lanka had the higher prevalence of CKDu, however, there was no increased risk posed by heat index, altitude, or land cover in multiple regression models. There was however an increased risk associated with strenuous agriculture cultivation practices. This suggests that physical exertion is a key factor in this region and is perhaps exacerbated by other unmeasured

confounders. In multiple regression models in India, vicinity to cropland (as a proxy for agrochemical exposure) posed a significant risk to eGFR<60, and like Sri Lanka, no increased risk was associated with heat index or altitude.

Linear mixed model analyses for both Sri Lanka and India showed very different results, as these models investigated effect of environmental variables on eGFR alone, whilst accounting for the fact that observations within the same country or zone tend to be correlated with each other. This is a factor which cannot be accounted for in linear and logistic regression, and thus could be a useful method to use when investigating environment-eGFR associations across CKD-endemic regions.

The issue of selecting the model which produces the "best fit" (i.e., a model with a low error) remains a vital area of data processing. Therefore, sensitivity analyses are invaluable to comparing the different types of interpolation methods which can be applied to the ERA-5 data, and in turn can help to reduce the risk of exposure misclassification when assigning heat index values to participants.

Although we must be cautious in our interpretation of these initial observations, these are a necessary and valuable starting point in the spatial and environmental characterization of CKDu and could aid surveillance and implementation efforts in future research.

5. A cross-sectional DEGREE study in urban and rural Malawi

5.1 Introduction

CKDu is an area of rapidly expanding research. However, efforts to determine the global and regional disease burdens, and investigate potential aetiologies, have been limited by variable study designs, inconsistent case definitions and investigations focussed on preconceived hypotheses of risk factors.

It is known that CKDu is affecting resource-constricted populations in Central America and South Asia and is mainly reported amongst young male agricultural workers in rural communities. However, standardized data are not available for comparison, meaning that the disease may have persisted for decades, and without a surveillance or treatment strategy in place the condition may be more widespread than previously thought.

The development of an internationally comparable standard is needed from which academic and clinical communities can estimate and compare disease prevalence and risk factors to improve research methodology across the field. The Disadvantaged Populations eGFR Epidemiology (DEGREE) Study [1] is an international collaboration led by PI Professor Neil Pearce (NP) at the London School of Hygiene and Tropical Medicine which has produced a standardized protocol to characterise the population distribution of the estimated glomerular filtration rate (eGFR) in low-and-middle income countries (LMICs) via a cross-sectional prevalence and risk factor study.

Although CKDu has been reported in Central America and South Asia, little information is available from African countries, which have similar climatic and occupational characteristics to reported CKDu-endemic regions, which would suggest that the rural population is likely at risk. Malawi was selected as a DEGREE study location as previous research in urban and rural Malawi has demonstrated a significant association between impaired kidney function and agricultural labour, which is a central hypothesis in the CKDu literature[32].

At the start of my PhD I began collaborating with NP on the Malawian DEGREE study as I was interested in the spatial an environmental aspect of CKDu, which had largely been overlooked in the existing literature. I designed my own nested environmental study (described in chapter 6) which was of interest to the DEGREE steering committee. As a result, I began collaborating with the Malawian DEGREE centre, which was beginning to survey in early 2018. It was agreed

between myself, NP and the Malawian PI Amelia Crampin that I would be granted access to the DEGREE dataset to use with my nested environmental study, and in addition to conduct descriptive analyses on the entire dataset for the wider team.

In this chapter, I will describe the DEGREE protocol cross-sectional prevalence study conducted amongst 821 adults in Malawi by the local clinical and field team. The study was conducted in an urban and rural site in 2018, collecting data on socioeconomic status, lifestyle and occupation covariates. Anthropometric measures and bio samples were also collected by the field teams. I will describe participant selection and data collection, which was conducted by the Malawian teams, and discuss the statistical analyses I implemented to the data below.

Ethical approval and funding

The DEGREE executive board was granted ethical approval by the Malawi National Health Sciences Research Committee protocol number #1912 (Lilongwe, Malawi), and the London School of Hygiene and Tropical Medicine Ethics Committee protocol #14709 (London, UK). The DEGREE study was funded by the Medical Research Council Global Challenges Research Fund Foundation Grant (MRC GCRF #MR/P02386X/1). See appendices 11a and 11b for ethics documents. Ethical approval from Imperial College was not required for this study is it is solely LSHTM-based.

5.2 Methods

5.2.1 Study design and objectives

The DEGREE survey aims to estimate the distribution of glomerular filtration (GFR) in an urban and rural population via a cross-sectional prevalence and risk factor study. This involves the quantification of renal function in a representative adult population-based sample and a requirement for standardization of serum creatinine (SCr) measurements, along with sample storage for cross centre validation of SCr values. Estimates of body composition were required to facilitate valid comparisons of estimated GFR (eGFR) within and between populations.

The specific study objectives were to:

(i) Obtain standardized estimates of eGFR in a rural and urban area

(ii) Determine the distribution of eGFR and estimate the prevalence of low eGFR (<60mL/min per 1.72m^2) in the two areas

(iii) Determine the proportion of those with low eGFR in whom history and clinical measurements are consistent with a diagnosis of CKDu

(iv) Assess risk factors associated with this estimated CKDu prevalence including age, gender, urban/rural residence, occupation and lifestyle factors

5.2.1.1 Study setting

Malawi is in Eastern Africa and has a predominantly agricultural-led economy with tobacco as its main export. At least 80% of the population is rural, but urbanisation is increasing annually by 4·1% due to high rural–urban migration and population growth [262].

The DEGREE study survey was conducted in Lilongwe, Malawi's capital city, and the rural Northern Karonga District (Figure 42). The study was nested in the existing Malawi Epidemiology and Intervention Research Unit (MEIRU) population-based survey of adults which was set up in 2013 to document the burden and drivers of NCDs in urban (Lilongwe) and rural (Karonga) Malawi, with the long-term aim of designing and testing intervention strategies [263]. This survey acted as the basis of the DEGREE sample frame, which is described below.

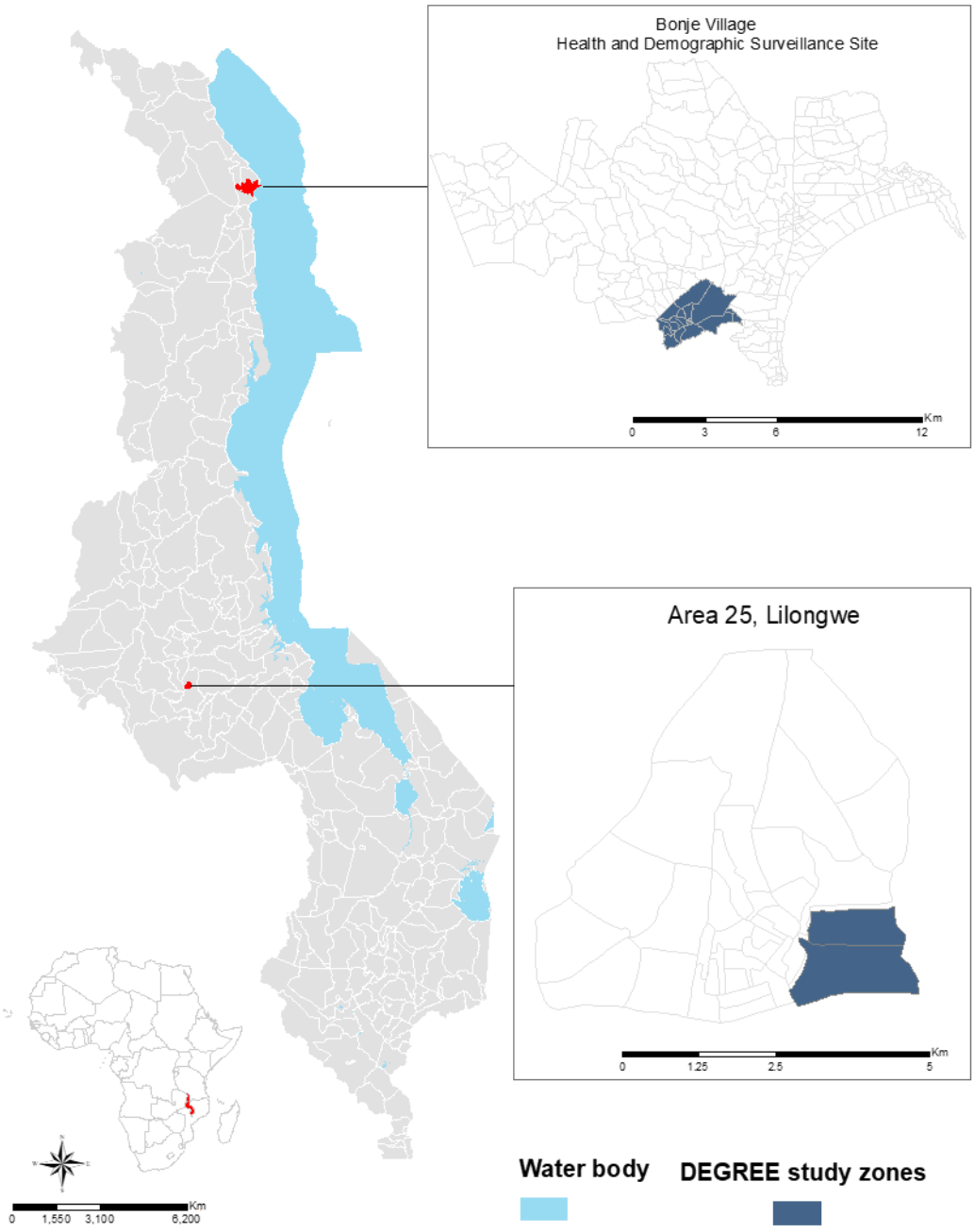


Figure 42. Map showing the urban and rural DEGREE study zones in Malawi

The rural DEGREE survey was conducted in the Karonga District in the well-established Health and Demographic Surveillance Site (HDSS) which was set up in 2002 with a baseline census to register all individuals and households in the area [264]. Vital events and migrations are reported by village informants, forming a continuous registration and demographic surveillance system. The HDSS is located between latitude -10.38° and longitude 34.08° covering 135km^2 with a

population of approximately 40,000 inhabitants [264], most of which are subsistence farmers. In the HDSS each village is divided into clusters of 35 households for vital reporting purposes [263], and groups of 10 clusters are combined to form 21 reporting groups, or study zones which I have shown in figure 43.

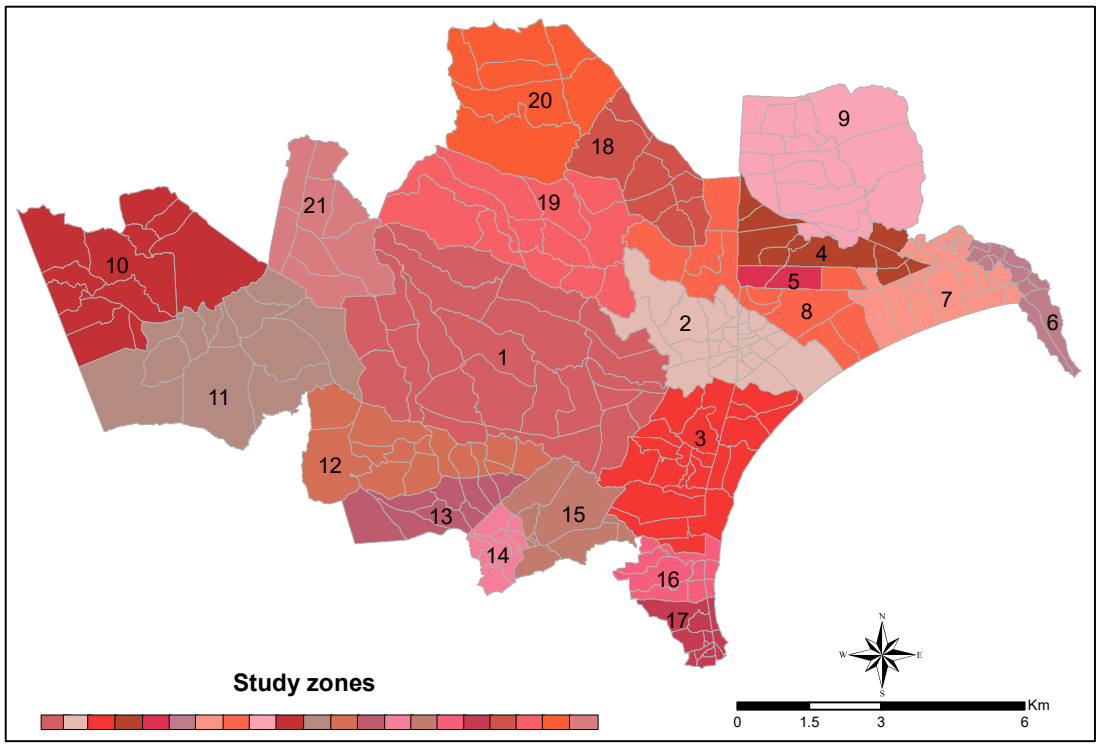


Figure 43. Map showing the Health and Demographic Surveillance Site (HDSS) study zones numbered 1-21, and colour coded to denote zone boundary

The urban DEGREE survey was conducted in Area 25, a sub-area of Malawi's capital city, Lilongwe. Area 25 is a high-density, economically mixed residential area where many casual and seasonal labourers reside. Located on latitude -13.95° and longitude 33.76°, Area 25 covers 10km^2 with an estimated population of 66,000[263]. Area 25 can be sub-divided into 35 enumeration areas (EA) which were demarcated by the National Statistical office in 2008. Each EA comprises approximately 480 households [263] which were used as the administrative boundaries of study zones for the DEGREE survey which I depict in figure 44.

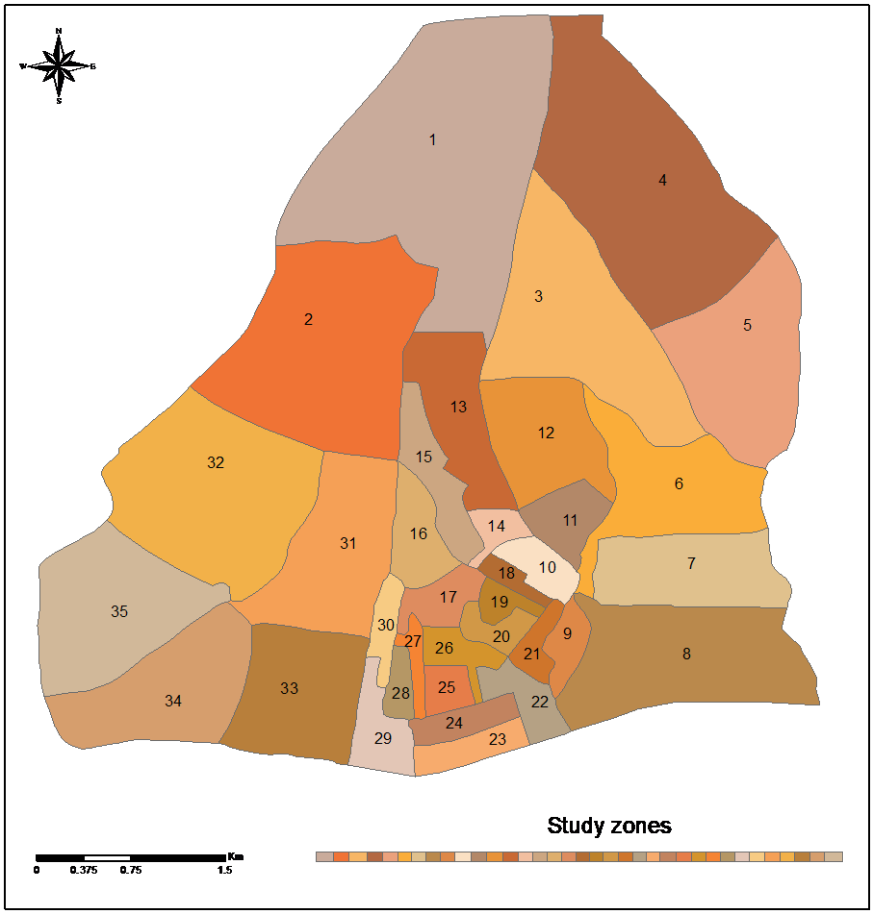


Figure 44. Map showing the Area 25 study zones 1-35, which are colour coded to denote zone boundary

5.2.1.2 Sampling strategy and participant recruitment

In the HDSS in Karonga, study zones 14 and 15 were sampled. These zones are located in Bonje (as will now be referred), an area in the south of the HDSS which surrounds a large rice irrigation scheme. In Bonje, 1130 participants ages 18 years and older were randomly sampled at the individual level from the existing HDSS NCD sample frame (figure 43 for reference).

In Area 25, study zones 7 and 8 in the South East part of Area 25 were sampled (figure 44). It was anticipated that many residents would have left since the 2013 MEIRU NCD survey due to seasonal employment outside the city. The field staff therefore re-enumerated study zones 7 and 8 and created a new list of residents as the sampling frame. This was done by dividing the existing enumeration zones which originally comprised 480 households, into smaller zones of approximately 100 households. After this, 1 in 3 adults were randomly sampled. Overall, 892

participants were randomly sampled in Area 25. In both study areas, the final sampling unit was at the household level.

To begin recruitment, field staff visited the selected households across Bonje and Area 25. If the participants were not at home on the first visit, staff would return twice more before excluding them from the recruitment process, coding them as "missing" in an electronic field log. Participants who had moved away or died since the 2013 MEIRU survey were coded as "Left" or "Died".

Those who were at home were given a participant information sheet outlining the study aims, procedures and data dissemination (see appendices 12 and 13 for English, Chichewa and Tumbuka language versions). These procedures were also explained verbally by field staff. Those who were willing to take part in the study signed a consent form (appendices 14, 15, and 16 for English, Chichewa and Tumbuka language versions) agreeing to be interviewed, and to opt in or out of having blood and urine samples taken. They were then given a unique pseudo anonymised study ID number (STID). All information and consent documents were designed by MEIRU and presented in English and regional languages Chichewa (Area 25) and Tumbuka (Bonje).

On the same day of consent, field staff asked permission to conduct data collection immediately to reduce loss between consent and interview. Prior to interview, the participant STID was added to an electronic study register, and the staff instructed participants who consented to give bio samples that they would need to fast overnight for blood and urine sample collection the following morning. As stipulated in the ethical approval this register was passed onto the clinical field team daily to identify which households to visit for bio sample collection. Figure 45 shows an overview of how the data was sampled across the study zones.

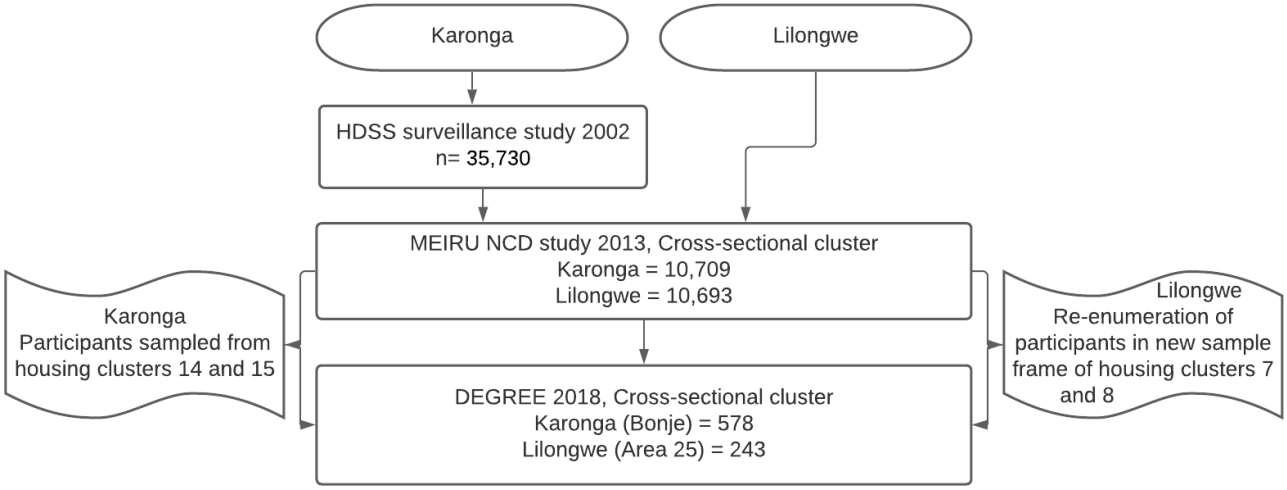


Figure 45. Flowchart showing the nested sample frames from which the DEGREE participants in Karonga (Bonje) and Lilongwe (Area 25) were sampled

5.2.1.3 Power calculation

A target sample size of 1500 male and female adult participants over 18 years of age was estimated to detect a mean eGFR 5mL/min (σ=30) difference between study sites (90% power; α=0.05). To detect a difference of 7mL/min in mean eGFR between two subgroups, a sample size of 250 adult participants in each group was required (80% power; α=0.05).

5.2.1.4 Data collection and laboratory analysis

Core questionnaire

Data collection for the urban and rural DEGREE survey were conducted by field staff between January and August 2018. Participants were interviewed using a core questionnaire based on the WHO Global Physical Activity Questionnaire [(265)] to obtain data on age, sex, income, schooling years, alcohol intake, smoking, meat consumption and medical history and occupational status. Anthropometric measurements including body mass index (BMI) and fat-free mass (FFM) were also taken. To ensure data integrity, questionnaire data was double entered into a field database using Open Data Kit (ODK) software (see appendix 17 for the core questionnaire).

Anthropometric measurements

Height and weight were measured twice per visit for each participant using calibrated Stadiometer seca-213, Scales seca-761 apparatus respectively (seca GmbH&Co. KG, Germany). The mean height and weight were calculated from these readings. Body composition was measured using a Bodystat 1500 bioimpedance analysis (BIA) instrument (Bodystat Ltd., United Kingdom). Fat-free mass (FFM) was calculated using the mean value of two bioimpedance output measures (fat-free mass/lean mass, kg). Body mass index (BMI, kg/m^2) was calculated by dividing mean body weight by mean body height and multiplying by 100. BMI was categorised as: (underweight: ≤18.5, normal weight: >18.5-≤25, overweight: >25-≤30, obese: >30). Blood pressure was measured with a portable electronic device OMRON HEM-7211-E, Model M6 (OMRON Healthcare Co., Ltd., Japan) on the right upper arm, three times in a seated position after 30 minutes of inactivity and with 5 minutes rest between measurements. Participants were classified as having hypertension if systolic blood pressure was ≥140mm Hg, or diastolic blood pressure was ≥90mm Hg, or if the participant was taking antihypertensive medication.

Urine samples

Clinical field staff collected one 30ml midstream fasting urine sample per participant, into a sterile test tube. Specimen tubes were labelled with a "SPECNO" label which listed participant sex, date of birth and STID. A specimen form was filled out detailing the procedure and type of sample taken.

Blood samples

Clinical field staff collected fasting venous blood samples using the evacuate method with a vacutainer tube. This method consists of a plastic hub, hypodermic needle and vacuum tube allowing blood to flow freely into a vacutainer tube. Three tubes (total volume of 18ml) were collected per participant for different laboratory procedures; 6ml ethylenediamine tetraacetic acid (EDTA) vacutainer tube for full blood count, 2ml Fluoride/Oxalate tube for glucose measurement, 10ml non-additive plain tubes for chemistry analysis and storage. To avoid cross contamination of additives between tubes, blood collection was conducted in the following

order: non additive, Fluoride oxalate, EDTA. Specimen tubes were labelled with a "SPECNO" label as described above.

Rapid HIV test

Malawi has one of the highest HIV prevalence rates in the world, with approximately 9.2% of adults (15-49 years) estimated as seropositive in 2018 [(266)]. During the interview, additional data was collected on HIV status as antiretroviral therapy (ART) – particularly nucleoside reverse inhibitors for HIV – are known to be nephrotoxic, and can therefore have an adverse effect on eGFR potentially overestimating CKDu prevalence if not accounted for[(267)]. Participants were asked about HIV status (if known) and any use of ART and were offered a rapid HIV test via a fingerstick test, collecting blood drops from the 3rd or 4th finger of the non-dominant hand. Blood drops were collected into a capillary tube and rotated and inverted to mix the blood in the testing device. Rapid tests results were reported to the participant immediately and reported in the core field database. For specimen collection and handling standard operating procedures please see appendix 18.

Chemical analyses of bio samples were conducted in regional MEIRU-run laboratories. Bonje samples were sent to the nearby Karonga Prevention Study (KPS) laboratory, and Area 25 samples were sent to the MEIRU headquarters laboratory in Lilongwe. For transportation detail and laboratory procedures please see appendices 19-22.

5.2.1.5 Ensuring data quality and data entry

All field staff were trained on the data collection protocol, including study description and taking informed consent. Data outputs recorded during fieldwork were reviewed by the fieldworker at the end of each day, and by the local PI at the end of each week to check for inconsistencies and/or missing data and would be sent back to a household if there was uncertainty or missing data from the questionnaires. Core questionnaire data was entered in duplicate, electronically using an Open Data Kit (ODK) on an encrypted tablet. Data were entered in an anonymized form using the participant STID. Laboratory results were uploaded into a centralised database located in the research headquarters in the MEIRU offices. All data was entered by field staff and laboratory managers in Malawi, and primary copies of all data is held by the local Malawian investigators. At the end of data collection, an anonymized copy of

the final dataset was transferred to the Data Hub at the London School of Hygiene and Tropical Medicine (LSHTM) in a password protected .csv file.

5.2.1.6 Data cleaning, coding, and analyses

After data collection finished in August 2018, the harmonized laboratory and questionnaire data which were collected and sampled by the Malawian field teams were made available to me via the Data Hub at LSHTM. As explained above, I was responsible for cleaning, coding and analysing this dataset, which I will explain below.

I cleaned the data by excluding participants with the missing basic variables date of birth, sex, and serum creatinine as these are key variables in the calculation of eGFR. Participants with missing glucose or blood pressure measurements were excluded, as without these variables I could not decipher the diabetic or hypertensive status of the participant, and thus whether a low eGFR was due to an underlying CKD-related co-morbidity, or due to suspected CKDu.

To avoid problems with small numbers in some categories, I re-grouped covariates with multiple categories into broader covariates. Based on the central hypothesis that CKDu is an agricultural disease, occupation status was re-grouped into discrete categories based on whether a participant was an agricultural worker or otherwise. Income and schooling were also re-grouped into tertiles ("unknown", "Malawian Kwacha 0 ≤20,000", "Malawian Kwacha >20,000"; and "≤5 years", ">5≤10 years", ">10 years" respectively). I also combined smoking and alcohol drinking to form a broad lifestyle covariate (See appendix 21, table 1 for original categories).

5.2.1.7 eGFR calculation

In sub-Saharan Africa (SSA), the true prevalence of CKD remains difficult to quantify [268, 269] largely due to poor sampling methods, and lack of study-specific standardization in creatinine measurement [191, 268, 270]. Further bias may also be introduced through the choice of the estimating equation for GFR. Guidelines developed by the National Kidney Foundation (NKF) recommend the use of creatinine-based equations such as the four-variable Modification of Diet in Renal Disease (4V-MDRD) to estimate GFR [271]. However this equation was based on a small study sample in North America with a coefficient that corrected for higher creatinine levels observed in African Americans [174, 191, 268], which is potentially inaccurate when estimating GFR in African populations [268, 272]. More recently, The Kidney Disease Improving

Global Outcomes (KDIGO) guideline recommended the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation as the preferred equation [50] which has a similar ethnicity-correcting coefficient for African Americans. However, when either of these equations with ethnicity-correcting coefficients were used in sub-Saharan African (SSA) populations, GFR was systematically overestimated [174, 272-275]. Some SSA studies have omitted the African American co-efficient in the absence of any appropriate and validated equations for African populations [276, 277]. DEGREE clinicians selected the CKD-EPI equation without the ethnicity-correction coefficient for the Malawian population, as this is the most recent standard practice guideline [50]. I calculated eGFR using the CKD-EPI equation shown below:

Equation 7. CKD-EPI equation to calculate eGFR

$$\text{eGFR}\ (\text{mL/min/1.73m}^2) = 141 \times \min(\text{S-Cr}/\kappa, 1)^{\alpha} \times \max(\text{S-Cr}/\kappa, 1)^{-1.209} \times 0.993^{\text{age}} \times 1.018\ [\text{if female}] \times 1.159\ [\text{if black}] \quad (49)$$

Where:

S-Cr is serum creatinine in µmol/L,

κ is 61.9 for females and 79.6 for males,

α is -0.329 for females and -0.411 for males,

min indicates the minimum of S-Cr/ κ or 1 and max indicates the maximum of S-Cr/ κ or 1

Defining CKDu: exclusion criteria

Clinical diagnosis of CKDu is based on two measures of serum creatinine three months apart [278] which is not practical in most epidemiological studies. As the DEGREE framework is mainly based in resource-constrained countries, the protocol specifies a diagnostic cut-off for reduced kidney function of eGFR<60mL/min/1.73m^2 in the absence of diabetes (fasting glucose >=7mg/l), hypertension (systolic ≥140 mm Hg, or diastolic ≥90 mm Hg), and heavy proteinuria (ACR>=30mg/mmol), based on a single measure [1].

In the event that there were few cases under this diagnostic status, I also investigated the prevalence and risk factors of reduced eGFR of <90 mL/min/1.73m^2 as this still signifies clinically significant kidney impairment in the absence of known CKD risk factors [279, 280].

5.2.1.8 Model building

The DEGREE protocol specifies a diagnostic cut off for impaired kidney function of eGFR<60ml/min per 1.73m^2 for reduced function in the absence of diabetes, hypertension, and heavy proteinuria[1]. There were few cases of impaired kidney function using this definition (see results below). I therefore decided to investigate risk factors for an eGFR below normal (<90ml/min per 1.73m^2), in the absence of known CKD risk factors.

I used linear regression models to estimate associations between postulated risk factors (age, sex, urban/rural residence, education, income, occupation, vegetarianism, lifestyle choices, and anthropometric measures BMI and FFM) and eGFR as a continuous outcome variable.

Logistic regression models were used to estimate associations between postulated risk factors listed above, and risk of eGFR <90 as a categorical outcome variable. For both linear and logistic regression, I included minimally adjusted models accounting for age sex and urban/rural residence, and fully adjusted models accounting for minimal adjustment covariates plus all other covariates (education, occupation status, income, lifestyle choices, vegetarianism and anthropometric measures BMI and FFM). In all models age was measured as a continuous variable in 10-year increments to estimate overall age-related association with continuous eGFR in the population. BMI and FFM were also measured as continuous variables in 5 kg/m^2 increments.

I observed a low seropositive HIV prevalence across the sample (0.3% (95% CI=0.07-1.0)); however, approximately 30% of the population had an "unknown" HIV status (n=223) which may have resulted from data entry error or participant reticence for testing due to social stigma of seropositivity to HIV. I conducted a sensitivity analysis to investigate if the "unknown" status HIV participants had a higher risk of reduced eGFR. Results showed no change in coefficients in either model, therefore the HIV variable was excluded from further models.

I conducted multiple regression modelling which included variables which were associated with reduced eGFR and risk of eGFR <90 in the minimally adjusted linear and logistic regression analyses (which were adjusted for age, sex and study site), along with variables of a priori interest identified via literature searches (study site, sex and education). I checked for multicollinearity each variable in the multiple regression model by comparing them to the covariates in the univariate models [281].

Furthermore, as CKDu is hypothesised as a predominantly rural disease, I conducted linear and logistic regression analyses for Area 25 and Bonje separately, with eGFR (continuous) and eGFR<90 (categorical) to investigate differences in risk factors (if any) between an urban and rural area. I conducted both minimally and fully adjusted models for each study site (as described above).

All analyses were conducted in R Studio version 3.5.1.

5.3 Results

5.3.1 Study population characteristics

The survey response rate was 56% (figure 46). When initially visited for recruitment by field staff, approximately 30% of these participants were classed as "missing" (n=559), and 11% refused to participate after reading the survey information documents (n=201). Thus, overall, 1076 participants out of 1908 initially selected, were eligible for the survey (n=654 females, n=422 males). After applying the a priori exclusion criteria, 821 participants were eligible for analysis (Figure 46). Table 38 summarizes the socio-demographic and anthropometric characteristics of the sample (see appendix 23, table 1 for characteristics of the non-restricted sample prior to exclusions, n=1076).

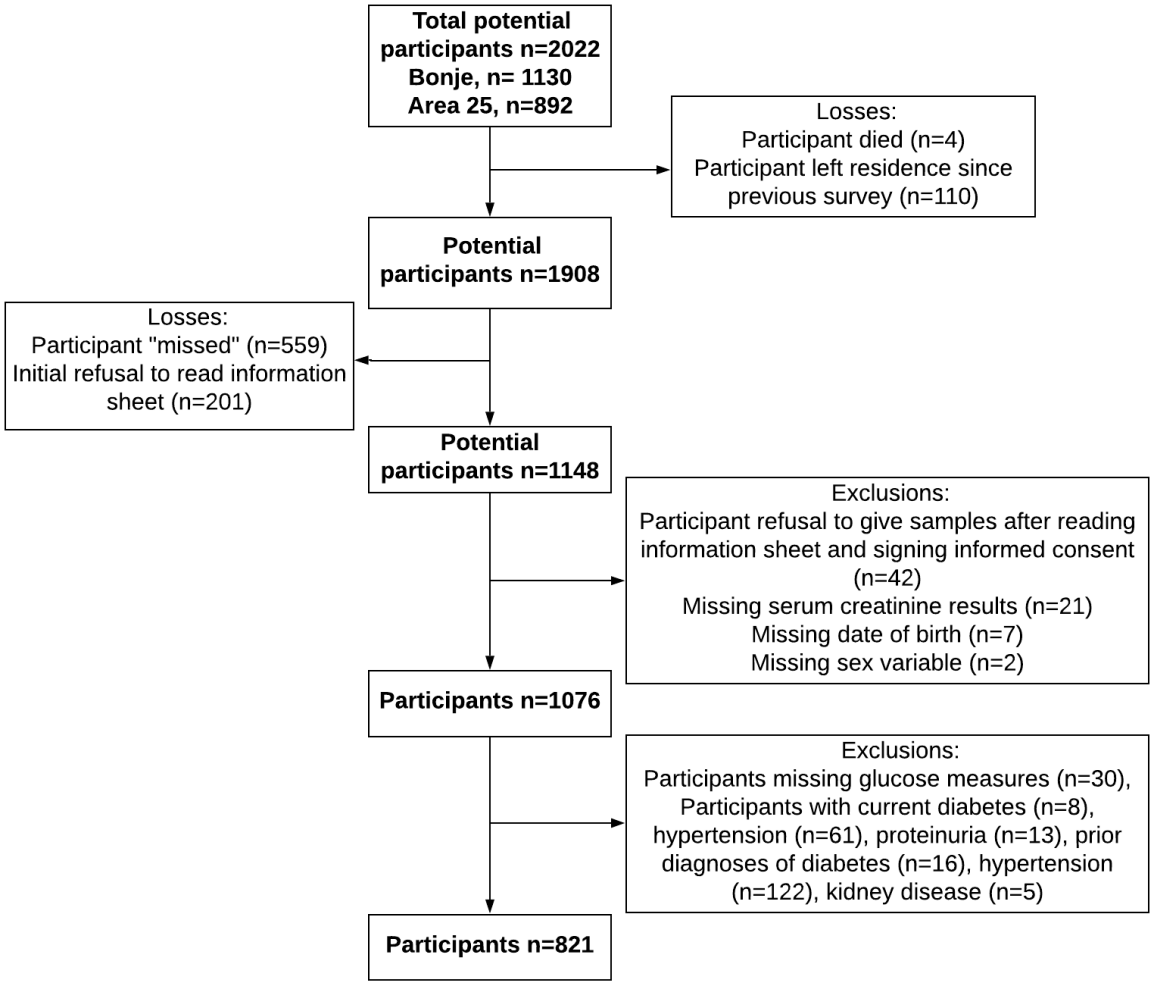


Figure 46. DEGREE exclusion criteria used for the Malawian sample frame, and final sample size used for analysis.

The mean (standard deviation (±SD)) participant age was 33.5±12.7 years. Mean BMI was 24.4±3.9 kg/m^2, and mean FFM was 51.7±82.3 kg/m^2. Mean systolic and diastolic blood pressure was 116.5±10.3 mm Hg, and 70.7±8.3 mm Hg respectively. Mean fasting plasma glucose was 47.7±0.7 mmol/l and the median (interquartile range (IQR)) albumin/creatinine ratio (ACR) was 0.5 (0.8) mg/mmol. Approximately 73% of the population were either tested for HIV during interview or provided information on their HIV status.

There was an 82% employment rate in the sample (females n=390 [77%], males n=286 [90%]). Approximately 57% of those in employment were classified as agricultural workers (female n=235 [61%], male n=148 [39%]). Overall, 56% of the population had completed >10 years of formal education.

Table 38. Sociodemographic and anthropometric characteristics of study participants without diabetes, hypertension and heavy proteinuria) n=821

Variable	eGFR n= 821		eGFR categories, n(%)[b] n= 2	n= 36	n= 783
	n(%)[a]	Mean (SD)	<60	>=60, <90	>=90
Age (years)					
<20	60(7)	132.6(13.2)	0	1 (2)	59 (98)
20-29	312 (38)	125.1(12.6)	0	2 (1)	310 (99)
30-39	232 (28)	115.7 (12.5)	1	4 (2)	227 (98)
40-49	130 (16)	107.8 (11.1)	1 (1)	2 (2)	127 (98)
50-59	46 (6)	100.9 (8.6)	0	6 (12)	40 (87)
60+	41 (5)	88.9 (9.9)	0	21 (51)	20 (49)
Sex					
Female	504 (61)	117.3 (15.6)	2 (0.4)	20 (4)	482 (96)
Male	317 (39)	116.81 (16.7)	0	16 (5)	301 (95)
Study site					
Urban (Area 25)	243 (29)	117.6 (14.9)	1 (0.4)	6 (3)	236 (97)
Rural (Bonje)	578 (70)	116.9 (16.5)	1 (0.2)	30 (5)	547 (95)
Education (years)					
≤5	59 (7)	109.1 (17.9)	0	9 (15)	50 (85)
>5≤10	301 (36	116.6 (16.4)	1 (0.2)	15 (3)	285 (62)
>10	461 (56)	118.5 (15.2)	1 (0.3)	12 (4)	448 (14)
Occupation					
Agricultural worker	383 (47)	113.9 (15.2)	1 (0.3)	20 (5)	362 (95)
Non-agricultural worker	438 (53)	119.90 (16.2)	1 (0.2)	16 (4)	421 (96)
Household monthly income (MK)[c]					
Unknown	17 (2)	118.6 (21.1)	1 (6)	0	16 (94)
MK 0 ≤20,000	406 (50)	117.8 (16.6)	0	20 (5)	386 (95)
MK >20,000	398 (49)	116.3 (15.1)	1 (0.3)	16 (4)	381 (96)
Healthy lifestyle choices					
Non-smoker/ never drink alcohol	630 (77)	117.0 (15.8)	2 (0.3)	28 (4)	600 (95)
Smoker/ alcohol drinker	191 (23)	117.3 (16.4)	0	8 (4)	183 (96)
Regular meat-eater					
Yes	621 (76)	117.5 (15.6)	2 (0.3)	25 (4)	594 (96)
No	200 (24)	115.6 (16.1)	0	11 (6)	189 (95)
Body mass index (kg/m^2)					
Underweight (≤18.5)	45 (6)	121.6 (18.0)	0	3 (7)	42 (93)
Normal (>18.5 - ≤ 25)	545 (66)	118.4 (15.4)	0	17 (3)	528 (97)

Overweight (>25 - ≤30)	177 (22)	113.3 (15.1)	2 (1)	7 (4)	168 (95)
Obese (>30)	54 (7)	112.1 (17.8)	0	9 (17)	45 (83)
Fat-free mass (kg)					
1st tertile (≤37)	124 (15)	115.1 (16.3)	0	8 (7)	116 (94)
2nd tertile (>37 - <45)	354 (43)	116.9 (16.9)	2 (1)	18 (5)	334 (94)
3rd tertile (≥45)	343 (42)	117.9 (14.9)	0	10(3)	333 (97)
HIV status					
Positive	3 (0.3)	114.3(15.1)	0	0	3 (100)
Negative	595 (73)	116.8 (15.1)	1 (0.2)	24 (4)	570 (96)
Unknown	223 (27)	118.0(18.1)	1 (0.4)	12 (5)	210 (94)

[a] percentage in columns; [b] percentage in rows; [c] Exchange rate (MK to USD) 0.001 at time of questionnaire

5.3.2 Mean eGFR and prevalence of reduced eGFR

The mean eGFR was 117.1±16.0 mL/min/1.73m^2. There was an inverse relationship between eGFR and increasing age and increasing BMI. Agricultural workers, those with <5 years of schooling, non-regular meat eaters (eating meat 0 days per week) and HIV positive participants also had a lower eGFR. Furthermore, those with earning a higher income (MK>20,000) also had a lower eGFR, which could be a result of higher earners having access to more unhealthy lifestyle choices such as a high fat diet.

A small site and sex-specific difference in mean eGFR was observed. The mean eGFR was 117.6(±14.9) for urban participants and 116.9(±16.5) for rural participants, and 116.8(±16.7) and 117.3(±15.6) for males and females respectively. Figure 47 shows the eGFR distribution of the sample population.

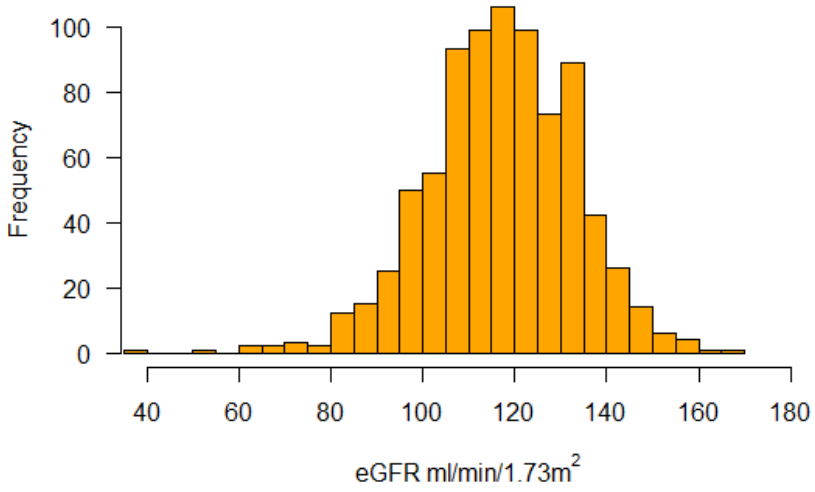


Figure 47. Histogram of eGFR distribution in the sample population across urban and rural Malawi

The prevalence of eGFR<60 in the study population was 0.2% (95% confidence interval (95% CI) 0.1-0.9). The prevalence of eGFR<90 was 5% (95% CI = 3.2-6.3). Site-specifically, there was a higher prevalence of eGFR<90 in Bonje (5% (95% CI = 3.6-7.8)) versus Area 25 (3% (95% CI= 1.4-6.7)). There was a marginal difference in prevalence of eGFR<90 between males (5% (95%CI=2.9, 8.0)) and females (4% (95%CI=2.7-6.5)).

5.3.3 Risk factors for reduced eGFR and eGFR <90

I first conducted linear and logistic regression models which included all potential risk factors and confounders; however, due to the small number of participants with an eGFR <90 (n=38), we also repeated the models minimally adjusted for key risk factors and confounders to test for inflation of relative risk due to sparse data. There was little difference between the findings from the fully adjusted and minimally adjusted models, therefore we report results from the fully adjusted models only. Table 39 shows linear and logistic regression models mutually adjusted for age, sex and study site (models 1 and 3), and models fully adjusted for all risk factor variables (age, sex, study site, BMI, FFM, education years, occupation status, lifestyle choices, income and non-regular meat eaters (models 2 and 4).

Model outcomes showed no significant differences between the eGFR coefficient and levels of adjustment for risk factors (10% change in regression coefficients), thus I report results from the fully adjusted linear and logistic regression models (2 and 4) only.

In the fully adjusted linear regression model (table 39, model 2), age was a key risk factor for reduced eGFR [regression coefficient (95% CI) -8.86 (-960, -8.15)] per ten-year age increase. Living in a rural site was independently associated with a higher eGFR [3.82 (1.51, 6.32)], despite this site having a higher prevalence of eGFR<90 (5%). No associations were observed between reduced eGFR and sex or occupational status.

Similarly, in the fully adjusted logistic regression model (table 39, model 4), the odds of having an eGFR<90 increased with age [OR= 3.59 (2.58, 5.21) per ten-year increase], and with increasing BMI [OR= 2.01 (1.27, 3.43) per 5kg/m^2 increase)]. There was no increased risk of eGFR<90 with sex, study site or occupation status.

Table 39. Associations between sociodemographic and anthropometric characteristics and eGFR and eGFR<90 in participants without diabetes, hypertension and heavy proteinuria, n=821

Variable	Model 1 Minimal adjustment	Model 2 Fully adjusted	Model 3 Minimal adjustment	Model 4 Fully adjusted
	eGFR Coefficient (95%CI)[a];	eGFR Coefficient (95%CI)[b];	eGFR <90 OR (95%CI)[a];	eGFR <90 OR (95%CI)[b];
Age [c] (per 10-year increase)	-8.73 (-9.37, -8.09)	-8.86 *** (-9.60, -8.15)	3.21 (2.49, 4.29)	3.59 *** (2.58, 5.21)
Sex [d]				
Male	0.02 (-1.62, 1.68)	-0.45 (-2.40, 1.47)	0.76 (0.34, 1.67)	1.44 (0.48, 5.05)
Female	Ref	Ref	Ref	Ref
Study site				
Urban (Area 25)	Ref	Ref	Ref	Ref
Rural (Bonje)	3.85 (2.05, 5.64)	3.82*** (1.51, 6.32)	0.69 (0.27, 1.92)	1.75 (0.50, 6.30)
Education (years)				
≤5	4.51 (1.10, 7.93)	4.29 (0.77, 7.69)	0.55 (0.14, 2.02)	0.64 (0.14, 2.56)
>5≤10	1.29 (-0.56, 3.15)	1.31 (-0.57, 3.18)	0.78 (0.30, 2.03)	0.86 (0.31, 2.39)
>10	Ref	Ref	Ref	Ref
Occupation				
Agricultural worker	-1.06 (-3.14, 1.01)	-1.11 (-3.20, 1.02)	0.44 (0.17, 1.21)	0.42 (0.15, 1.17)

Non-agricultural worker	Ref	Ref	Ref	Ref
Household monthly income (MK) [e]				
Unknown	-1.86 (-7.60, 3.87)	-2.26 (-8.01, 3.50)	1.39 (0.06, 13.66)	2.21 (0.09, 20.58)
MK 0 ≤20,000	Ref	Ref	Ref	Ref
MK >20,000	-0.34 (-2.16, 1.47)	-0.01 (-1.85, 1.87)	1.32 (0.54, 3.23)	1.05 (0.40, 2.78)
BMI (kg/m^2)				
5kg/m^2 increase	-0.99 (-2.8, 0.09)	-0.94 (-2.02, 0.17)	2.00 (1.29 3.15)	2.01 ** (1.27, 3.43)
Fat Free Mass (kg)				
5 kg/m^2 unit increase	-0.02 (-0.07, 0.02)	-0.02 (-0.07, 0.02)	1.00 (0.00, 1.02)	0.99 (0.003, 1.03)
Healthy lifestyle choices				
Non-smoker or alcohol drinker	-1.49 (-3.57, 0.59)	-1.43 (-3.50, 0.70)	2.07 (0.72, 6.64)	2.06 (0.68, 7.06)
Smoker and alcohol drinker	Ref	Ref	Ref	Ref
Regular meat-eater				
Yes	-0.40 (-2.32, 1.51)	-0.33 (-2.27, 1.65)	1.34 (0.58, 3.39)	1.04 (0.43, 2.72)
No	Ref	Ref	Ref	Ref

[a] minimal adjustment for age, sex and area, [b] all variables mutually adjusted; [c]adjusted for sex and area, [d]adjusted for age and area; [e] Exchange rate (MK to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

5.3.4 Urban/rural comparison of eGFR

Fully adjusted linear and logistic regression results are reported for Bonje (Table 40, models 1 and 2). Due to the small proportion of participants in the eGFR<90 category in Area 25, I could not conduct logistic regression analysis for this study site thus, fully adjusted linear regression results are reported (Table 40, model 3). For minimally adjusted models refer to appendices 24 and 25.

In both Bonje and Area 25, increasing age was associated with reduced eGFR (Table 40, models 1-3), and in Bonje there was a significant increased risk of reduced eGFR with age [OR = 3.09 (2.12, 4.59)] (model 2). In linear regression model 1 for Bonje, education <5 years was associated with a higher eGFR [regression coefficient (95%CI) 4.19 (0.39, 7.99)], even though

this category has the lowest mean eGFR (108.6mL/min/1.73m^2). After further stratification by age group, the positive association between education <5 years and eGFR did not remain [regression coefficient (95%CI) 3.28 (-0.61, 7.17)]. This suggests that this association was confounded by age and higher proportion of participants in this education category having an eGFR >90. In this category, eight participants had an eGFR<90 (17%), all of which were aged 60 years plus. The remainder of the participants in this category had an eGFR>90 (83%), thus creating a protective effect for those with the lowest mean eGFR and the fewest years of schooling.

In logistic regression models for Bonje, increasing BMI was associated with risk of eGFR<90 (models 1-2). No further associations between eGFR and risk factors were observed at either study site.

Table 40. Associations between sociodemographic and anthropometric characteristics and estimated glomerular filtration rate (eGFR) (fully adjusted) after removal of those with hypertension, diabetes and proteinuria, Area 25 (n=243) and Bonje (n=578)

Variable	Bonje Model 1 eGFR Fully adjusted Coefficient (95%CI);	Bonje Model 2 eGFR <90 Fully adjusted Coefficient (95%CI)[a];	Area 25 Model 3 eGFR Fully adjusted Coefficient (95%CI)[a];
Age			
Per 10-year increase	-9.00***	3.09***	-8.14
	(-9.79, -8.22)	(2.12, 4.59)	(-10.61, -6.23)
Sex			
Male	-0.10	1.55	-1.62
	(-2.45, 2.23)	(0.37, 8.28)	(-5.29, 2.05)
Female	Ref	Ref	Ref
Education (years)			
≤5	4.19**	0.92 **	5.59
	(0.39, 7.99)	(0.17, 4.58)	(-2.63, 13.82)
>5≤10	1.65	1.04	-0.62
	(-0.38, 3.69)	(0.33, 3.40)	(-5.49, 4.24)
>10	Ref	Ref	Ref
Occupation			
Agricultural worker	-1.04	0.45	0.99
	(-3.14, 1.04)	(0.17, 1.27)	(-13.76, 15.75)
Non-agricultural worker	Ref	Ref	Ref
Household monthly income (MK)[b]			
Unknown	-4.46	3.18	-0.83

	(-11.76, 2.82)	(0.11, 37.04)	(-11.10, 9.43)
MK 0 ≤20,000	Ref	Ref	Ref
MK >20,000	0.35	1.02	-1.73
	(-1.67, 2.37)	(0.37, 2.73)	(-6.45, 2.98)
BMI (kg/m^2)			
5kg/m^2 increase	-0.62	1.89**	-1.58
	(-1.97, 0.73)	(1.06, 3.81)	(-3.54, 0.36)
Fat Free Mass (kg)			
(Per 5 kg/m2 unit increase)	-0.01	0.92	-0.02
	(-0.07, 0.04)	(0.58, 1.01)	(-0.11, 0.06)
Healthy lifestyle choices			
Non-smoker or alcohol drinker	-1.40	1.52	-1.84
	(-3.95, 1.13)	(0.43, 5.87)	(-5.83, 2.14)
Smoker and alcohol drinker	Ref	Ref	Ref
Regular meat-eater			
Yes	-0.10	1.05	-1.95
	(-1.77, 2.77)	(0.41, 2.93)	(-7.38, 3.46)
No	Ref	Ref	Ref

[a] all variables mutually adjusted; [b] Exchange rate (MK to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

5.4 Discussion

The distribution of eGFR across the population sample showed a high proportion of participants with an eGFR >120 ml/min/1.73 m^2, suggesting that low eGFR is not generally a health burden in this population. Only two cases of eGFR<60 were observed, so I therefore adjusted the reduced eGFR cut-off value to eGFR<90. Overall, reduced eGFR and the risk of reduced eGFR were associated with increasing age; and increasing BMI was associated with an increased risk of reduced eGFR. These risk factors are commonly associated with reduced kidney function, both with and without diabetes, hypertension or proteinuria [172, 282].

In site-specific analysis, age was associated with reduced eGFR in both linear and logistic regression models. Furthermore, in Bonje increasing BMI was a risk factor for reduced eGFR. No other associations were observed in Area 25.

Contrary to the current literature that CKDu is likely a rural disease [as discussed in chapter 2] [2, 182, 283], there was a positive association between rural study site and eGFR after adjustment for risk factors, despite the higher crude prevalence rate observed here. This could be due to the higher proportion of older inhabitants in Bonje which, together with the well-established association between age and renal decline, explains the higher prevalence of eGFR<90 (5%) that was observed. This lower risk observed in Bonje could also be further inflated due to

unmeasured confounders in the area such as higher physical activity levels and lower ambient pollution levels. Research into the causal relationship between air pollution and the kidney damage is in its infancy, however there is increasing evidence suggesting that high ambient air pollution levels (specifically particulate matter) can harm the kidneys via oxidative stress resulting in albuminuria and a lower eGFR [284, 285]. In the case of CKDu this relationship is yet to be investigated, however this could be a significant area for future research.

Causality concerning the environmental risk factors of CKDu is debated across the literature, as, despite intensive research and debate, the specific aetiological agent causing CKDu remains elusive. The issues with chronic kidney diseases are that they are frequently multi-causal and largely require longer periods of time and more rigorous research to disentangle. In a recent study the complexity of the potential causal risk factor Glyphosate was mapped out using a "web of causation"[22] followed by a potential mechanism of this pesticide toxicity on the kidney. This research however was speculative and did not incorporate bio samples into the analyses and therefore is more speculative of causality. This is similar for many other areas of current CKDu research, whereby association analyses are conducted which serve as a basis for causal hypotheses which require further, more rigorous study designs and analyses to determine[10, 13, 71, 76, 286, 287]. Thus far, as CKDu research is in its infancy, many causal assertions of environmental exposures have been based on ecological assumptions and therefore strong evidence to causal mechanisms are weak; however, establishing causality in epidemiology often takes decades of exhaustive research using larger population samples, appropriate bio samples and markers for disease. In the case of this Malawian population, it is unlikely that genetics play a role in having a low eGFR.

I did not observe a high prevalence of eGFR<90 in the working age population in Bonje. In this village, rice is the predominant crop and I assumed most of the agricultural workers from the rural sample were employed in the neighbouring schemes. However, no increased risk was observed in this occupational (and age) group. This finding is contrary to the literature that agricultural workers in Central America and rice workers in South Asia who conduct similar levels of strenuous labour in similar climatic conditions have an increased risk of low eGFR [13, 69, 164, 214]. This observation from the Malawian sample could count against the current heat hypothesis which is central in the research landscape [11, 31, 288, 289].

Strengths of the study include the random selection of population-based participants, and stringent measurement protocols used in the field. This is also one of the first CKD-based studies conducted in Malawi, and the large amount of data collected in this population will inform future NCD studies in the country. Finally, DEGREE definitions of CKDu were used, and standardized laboratory measurements were adhered to. This will facilitate international comparisons of CKDu prevalence and help describe risk factors and identify the causes and mechanisms leading to the disease.

The study also has some limitations. Firstly, this is a cross-sectional study, which meant only one eGFR measure was taken and thus I could not differentiate acute kidney injury (AKI) from CKD. I therefore may have misclassified some AKI as reduced eGFR and overestimated the prevalence of reduced eGFR (<90). There is debate surrounding the diagnostic criteria for CKDu. This is mainly due to the exclusion of hypertension and proteinuria, both of which are both causes and consequences of advanced CKD (and CKDu). However, at early stages, proteinuria is uncommon in CKDu, and hypertension is not present, by definition; however, both are common in other forms of CKD (including diabetic glomerular disease), and at the population level, most disease is likely to be detected at earlier stages. The rationale for the 'DEGREE definition' (of eGFR<60 as a proxy for CKDu) is therefore that although this may result in misclassification of a few cases of advanced CKDu as non-CKDu, this is likely to be rare, whereas not doing such exclusions would lead to a large overestimation of CKDu prevalence.

Secondly, as this population was nested in an existing survey, selection bias could have occurred, potentially attenuating estimates. In the HDSS participants were easy to locate and had a 70% response rate from initial visit. In Area 25 however the response rate was 45% which was likely due to external employment in casual agricultural labouring. As a result of differential response rates between study sites, statistical power was limited which does not allow us to classify this study as having a negative result but possibly a "non-positive" result. This is reflected in the wide confidence intervals – specifically in age which is a common risk factor for CKD- in linear and logistic regression models. It would have been valuable to conduct participation bias analyses to understand which sub-group were predominantly refusing to take part in the study, however there was no baseline sociodemographic data available to me from those who refused, nor those from the previous studies in the area from which the nested studies were sampled.

Furthermore, the CKD-EPI equation has not yet been validated for the African population, and it is possible that calculated eGFR is not fully reflecting underlying kidney function accurately in the population. Finally, as with all epidemiological studies, there may have been some residual confounding that was not adjusted for, such as environmental factors, childhood illness or infection which could have pre-disposed certain participants to have a lower eGFR, and other socio-economic and lifestyle factors.

5.4.1 Conclusions

In conclusion, these findings indicate that in Malawi reduced eGFR consistent with the definition of CKDu is very unlikely to be a public health issue, and that the absence of site, occupation and sex associations, CKDu is not present in the Malawian population, at least to the extent that it has been observed in other tropical and sub-tropical countries in Central America and South Asia. This population had a relatively normal eGFR distribution, with a higher mean eGFR than CKDu endemic regions. As a result of so few participants having an eGFR <60, those captured with a lower eGFR (eGFR<90 in this sample population) are potentially pre-diabetic or hypertensive, and thus will eventually develop "traditional" CKD. These results are informative to the CKDu research field as the disease was not observed in a region which has a very similar climate and agricultural-led economy to reported CKDu-endemic regions. This "negative" finding is important as it contradicts some current hypothesis that CKDu is endemic across tropical and sub-tropical countries and supports the contention that the causes of the epidemic(s) remain unknown and more evidence is needed[143, 288].

6. Nested environmental study, Malawi

6.1 Introduction

Results reported in chapter 5 indicate that reduced eGFR consistent with the definition of CKDu is non-existent in the Malawian population. An eGFR<90 in the absence of known CKD co-morbidities was present, but these cases were associated with traditional CKD risk factors age, and BMI. Furthermore, contrary to the current literature that CKDu is a rural disease which disproportionately affects males, there was low prevalence of reduced eGFR amongst males, and no rural-urban associations with reduced eGFR were found.

These interesting findings suggest that risk factors for CKDu are not homogenous and poses the question as to why there are no cases of reduced eGFR or CKDu amongst the Malawian population who work in similar agricultural occupations and in similar climatic conditions as those CKDu affected countries. These observations warrant more detailed analyses on potential environmental and lifestyle risk factor to elucidate which (if any) of the widely hypothesised risk factors might be associated with a lower eGFR in Malawi. These results can in turn facilitate direct international comparison between DEGREE sites.

As discussed in chapter 5, the core function of the DEGREE survey is to estimate the distribution of eGFR using standardized methods to determine the proportion of those with measurements consistent with the diagnosis of CKDu[2]. The scope of this survey, however, does not include hypothesis testing of environmental risk factors which feature heavily in the CKDu literature. To address this issue, I designed a nested environmental study aiming to capture detailed environmental and lifestyle data, which, coupled with core DEGREE data will enable investigation of the associations between environmental exposures and low eGFR at the individual-level.

In chapter 2 I highlighted the key environmental variables which are hypothesised as risk factors for CKDu but are generally not investigated in depth. I designed this nested environmental study to investigate the four facets which are highlighted below:

(iv) Conducting an extensive environmental questionnaire to gather self-reported information on exposure to potential occupational, lifestyle and environmental risk factors for CKDu

(v) Measuring biomarkers of heavy metals exposure in samples of whole blood

(vi) Quantifying heavy metal concentrations in local potable water sources

(vii) Analysing the geographical risk factors land cover type (as a proxy for exposure to pesticides), altitude, and heat index in relation to participant residence using satellite imagery and climate reanalysis data

I conducted a nested cross-sectional study among 367 adults across Bonje (n=258) and Area 25 (n=109) and explored the effects of environmental, occupational and lifestyle risk factors on eGFR. As results from chapter 5 show that this population do not have a high prevalence of CKD, risk factors are modelled with eGFR as a continuous variable.

Using an extensive questionnaire, participants were asked to self-report on exposure to environmental risk factors including farming practices, use of pesticides, and potable water sources. Furthermore, biomarkers of exposure to heavy metals in biological samples were measured, along with the concentration of heavy metals in potable water sources. Geographical, and meteorological characteristics derived from satellite data were also assessed, specifically, residential vicinity to land cover type, altitude, and heat index values. I discuss participant selection, data collection and analysis for each of the environmental study facets in the following order: (i) questionnaire data and bio sample collection, (ii) potable water collection (iii), environmental satellite data assignment to participants.

Ethical approval was granted by the National Health Sciences Research Committee Protocol # 17/09/1912. This nested study was funded by the Medical Research Council Global Challenges Research Fund Foundation Grant (MRC GCRF #MR/P02386X/1) (appendix 26); and the MRC Supplement Scheme #63.

6.2 Methods

6.2.1 Environmental study part 1: questionnaire and bio sample collection

6.2.1.1 Study design overview

This nested study was conducted in Bonje and Area 25 to investigate the association between environmental exposures and low eGFR at the individual level. The study was conducted on a random sub-sample of the wider DEGREE population described in chapter 5. Participants were

administered a detailed questionnaire, and bio sample aliquots were taken to measure heavy metals concentrations. Water samples were taken from each study zone using a random stratified sampling technique and were measured for a broad range of heavy metals. Participants were assigned heavy metals concentrations from the nearest water sources to their residence. Finally, similar to chapter 4, environmental data derived from satellite imagery was processed and assigned to each participant. These data were then merged into a harmonized dataset and analysed.

6.2.1.2 Sampling strategy and participant recruitment

The DEGREE field team invited every second household they visited daily when conducting the DEGREE survey data collection to take part in the study. Along with the core DEGREE information and consent form, participants were also given an information sheet which I created, outlining the nested study aims, procedures and data dissemination plan (appendix 27). These procedures were also explained verbally by field staff. Those willing to take part in the nested study signed the consent form (appendix 28) agreeing to be interviewed and to have their bio samples tested for heavy metals.

6.2.1.3 Power calculation

The overarching DEGREE study is powered at 80% to detect a difference of 7mL/min eGFR between populations, with a SD of 30ml between regions, with a sample size of n=1000 per study site (α: 0.05, β-1: 0.8, σ: 30, δ:7). Under this scenario 4ml/min was the smallest possible difference that could be detected between populations. Using this prerequisite (α: 0.05, β-1: 0.8, σ: 30, δ:7) I calculated the sample size for the nested study. A target sample size of 289 male and female adult participants over 18 years of age per study zone was estimated to detect a mean eGFR 7mL/min (σ=30) difference between study sub-groups (80% power, α=0.05).

6.2.1.4 Environmental questionnaire - development and training

Together with a colleague at IS Global in Barcelona, I designed a questionnaire which collected data on lifestyle factors and environmental variables. The questionnaire was designed to accompany the core DEGREE survey (see chapter 5) and was administered at the end of core data collection in sampling households. The questionnaire contained structured questions which I developed based on earlier research of environmental risk factors of CKDu in other areas of the world (see chapter 4), as well as input from the Malawian field work teams. The data collection tools were designed to capture data on site-specific environmental and lifestyle

risk factors for CKDu including detailed occupational data, farming practices, and drinking water sanitation methods which to date lack detailed investigation. Similar to the core questionnaire, the EMQ was broadly based on the WHO STEPS questionnaire[265], and comprised four main sections which are detailed in table 41.

Table 41. Summary of variables collected in the environmental module questionnaire as art of data collection for the nested study

Section	Information gathering	Variables collected
1	Household information	Time lived in residence
2	Occupation information	Occupational category Time worked in occupation(s) Daily working hours and rest break duration Physical symptoms resulting from occupation(s)
3	Agricultural practices	Crops cultivated and pesticide use practices Use of personal protective equipment when applying pesticides
4	Lifestyle information	Diet Drinking water sources Water treatment practices

The occupation section of the questionnaire aimed to ascertain whether a participant was engaged in physically straining employment, as this is thought to be a major risk factor for low eGFR and CKDu [14, 16, 17, 289]. This section was particularly challenging to design as the labour market in Malawi is largely characterized by casual and seasonal employment, with a high rate of internal migration[290, 291]. A further challenge was capturing the activities of those who are in unpaid domestic work, as in the Malawian setting, this usually consists of strenuous daily activities including land tilling and subsistence farming, all of which can cause kidney damage. After discussion with l field workers and MEIRU colleagues about the sectors of Malawian employment, occupation was categorized as follows: “street trader/vendor”, “shop trader/vendor”, “manufacturer”, “office worker”, “construction worker”, “commercial agricultural worker”, “casual labourer”, “subsistence farmer/fisher”, “carrying or transporting items/people”, “homemaker (household/land activities)”, “full time student”, and “other”. In addition to this, participants were asked to report on working hours, rest breaks and physical

symptoms experienced during their occupation or daily activities. The symptoms chosen were those which are well documented in the CKDu and wider medical literature as synonymous with heat stress, and dehydration and thus potentially detrimental to eGFR[16, 17, 31, 71].

The agricultural section of the questionnaire was designed to capture the farming practices of participants tending crops from the commercial to subsistence scale. Crop type and plot size were important variables as they signalled the likely intensity of labour required for crop upkeep. The crop type and plot size variables were selected based on the literature[292, 293], and from discussion with MEIRU colleagues. Plot size was categorized into quintiles from <1 hectare to >15 hectares. For the full list of crop types see appendix 29. As part of the agricultural section, there was also a detailed portion on pesticide handling practices, which was largely modelled on a previous questionnaire conducted in Ethiopia [294].

At the time of data collection, measurement of pesticides in potable water and bio samples was included in the scope, however, due to a lack of reliable standards at the time in both the UK and Malawi, this analysis path was excluded. Nonetheless, questionnaire information from the pesticide practices section was kept in the main analyses as most farmers in Malawi use pesticides [295], and inferences could be gleaned from modelling these self-reported data with eGFR. Data was collected on pesticide mixing, application, and storage practices, along with use of PPE, and for each question, participants could select more than one option. There was a free text section for participants to list the names of the pesticides or the known active ingredients used.

The final section of the questionnaire was an extension of the core DEGREE questionnaire and was designed to capture more detailed lifestyle data, mainly based around diet, drinking water sources, and water sanitization treatments. As discussed in chapter 2, there is a growing body of research on exposure to heavy metals from potable water [23, 24, 30, 61, 296, 297]. Drinking water is often considered to be a source of nephrotoxic contaminants, and anecdotal evidence suggest that those consuming spring water have a higher eGFR than those drinking from wells and other groundwater sources [79].

In Malawi there is sparse information on potable water sources, so I finalized this section of the questionnaire after I conducted reconnaissance visits around the study zones in March 2018. I was able to visit a broad range of potable water sources across both study sites and

learned that sources are generally used on a seasonal basis, depending on the height of the water table over rainy and non-rainy seasons. Through researching the literature, and conducting informal interviews with residents, I compiled a list of the most reported sources. In the questionnaire, there are 18 water sources listed, and whether the source was primary (all year use) or secondary (seasonal use). Information on daily water intake was also collected and was based on a standard plastic cup measure (500ml). Water sanitization practices were classified into the most commonly used in the country: "boiling", or "WaterGuard" (chlorine).

In the final section of the questionnaire, data was collected on a participant average weekly diet. Diet is known to have an impact on renal function, particularly with high consumption of protein. Dietary questions were categorized by foodstuff: meat, fish, pulses, rice, cassava, maize, offal, vegetables; and quintiles of consumption per week: 1-2 times, 2-3 times, 3-4 times, >4 times and never.

Questions were largely closed ended and consisted of short sentences, and I included explanatory text above each question to guide the interviewer. When conducting studies, there is tension between the desire to reduce costs and increase the informativeness of surveys[298]. As this questionnaire was exhaustive, I assumed that some sections might not be applicable to all participants, such as details on farming methods amongst some urban residents. To maximise efficiency, I therefore included skip sequencing, in which the response to an opening question is used to determine whether a respondent should be asked certain subsequent questions[299]. The objective is to eliminate inapplicable questions, thereby reducing interview time. Due to this being a pilot study, it was decided between myself and MEIRU that the questionnaire would be in a paper format so as not to delay the use of the existing ODK system for the field workers. As this format did not support electronic skips, I made the navigational instructions for the interviewer as clear as possible by placing skips immediately after an answer that lead to the branch point in the questionnaire. Skip instructions were worded positively (e.g., "Go to question 2.1") as opposed to negatively ("Skip question 3.2").

The questionnaire was translated into the local languages Tumbuka (Area 25) and Chichewa (Bonje), and back translated into English by the local DEGREE administrative teams in each study zone. Before the nested study commenced, I held afternoon training sessions with both the Bonje and Area 25 field teams, explaining the aims and objectives of the questionnaire,

and how to record information. We simulated mock interviews together to ensure that the translation and back translations were correct, to amend for any issues in the interview process. I also timed the interviews to allow field workers to plan their daily timetables efficiently.

6.2.1.5 Data collection and data quality

Interviews commenced in Bonje and Area 25 in March and April 2018 respectively and finished in May. The interviews were conducted immediately after the DEGREE core questionnaire and took approximately 20 minutes to complete. At the end of each day, paper questionnaires were brought back the local offices near Bonje and Area 25 and were reviewed by both the fieldworker, and by the field team lead to check for inconsistencies and/or missing data. The environmental questionnaire data was collected in an anonymized form using the participant STID (as described in chapter 5). After data collection was complete the questionnaires were scanned into an online shared database.

6.2.1.6 Biological matrix to measure heavy metals

As part of the core DEGREE study, blood and urine samples were collected from participants (detailed in chapter 5) to measure eGFR. For the purpose of the nested environmental study, a 2ml aliquot of those blood and urine samples were collected from consenting participants to be analysed for heavy metals.

Due to budget constraints, I decided to have one of the two biological matrices (blood or urine) tested first to assess whether high concentrations of heavy metals were observed before testing the second matrix.

Across the literature, hair analysis has been understood to play an important role in heavy metals monitoring. Human hair can accumulate heavy metals and can incorporate metals into its structure during growth, and therefore can reflect the mean level in the human body[300]. Furthermore, compared to the other matrices such as blood and urine, hair has advantages, easy collection and transportation, and reflects long-term exposure to heavy metals than in other matrices[301]. Although this would have been the overall optimal matrix, collection of hair samples and subsequent analyses were not in the scope of the ethics for the overarching DEGREE protocol and I therefore could not use this as a bio sample for measurement.

Whole blood is also a widely used matrix for heavy metals testing, specifically in testing for lead exposure[302]; however, it is also a very complex matrix, due to clotting potential in the freeze-

thaw process before analysis. Urine is a widely used matrix for heavy metal testing and is easily sampled; however it is one of the more difficult biological matrices to analyse due to its high concentration of inorganic salts which can cause interference in the determination of concentrations and affect accuracy of the analysis[303]. Furthermore, urate salts can precipitate out of urine during transport and storage and can be a complicating factor for analysis, particularly if the samples are shipped internationally [304].

These factors were considered carefully, and whole blood was selected as the first matrix to test, as, from reviewing the literature, and discussions with the laboratory teams in Malawi and the UK, this was optimal matrix for a majority of the metals of interest.

There was concern with Arsenic and Cadmium however, as for these heavy metals, urine is typically the optimal matrix for detection. In the case of Cadmium, blood concentrations reflect acute exposure whilst urine concentration more closely reflects total body burden. However, it has been suggested that when renal damage from Cadmium exposure occurs, the excretion rate decreases, and urinary Cadmium concentrations no longer reflect body burden[305].

In the case of Arsenic, blood is a less sensitive biomarker of exposure due to its short half-life. However, if exposure is continuous and steady, the blood Arsenic levels can reach a steady-state and reflect the degree of exposure[306].

The blood samples were aliquoted in the Malawian laboratories, they were stored in ethylenediamine tetraacetic acid (EDTA) vacutainer tubes at -20^{o}C until collection was complete in May 2018. Specimen tubes were labelled with a "SPECNO" label as described in chapter 5.

6.2.1.7 Transport and laboratory analysis of bio samples

The aliquots were shipped on dry ice to the University of Nottingham for heavy metals analysis (for the Materials Transfer Agreement (MTA) for shipping bio samples, see appendix 30). Laboratory staff checked the condition and volume of the samples and reported any issues in an Excel database. Multi-elemental analysis of whole blood by ICP-MS, run with a dilution of 0.3 mL in 10 mL diluent (0.5 % Primar Plus grade Nitric acid). For full description see appendix 31.

6.2.2 Environmental study part 2: Potable water collection

6.2.2.1 Study design

Between March and April 2018 (Malawi's rainy season), I collected potable water samples from Bonje and Area 25 using a random sampling technique stratified by water source. As discussed in chapter 5, the Karonga HDSS is a well-established surveillance site, which had a geocoded map of major water sources across the area. In addition to shallow hand dug potholes, the major drinking water sources in this area were boreholes and wells (Figure 48a and figure 48b). Geocoded maps were not available for Area 25, so I conducted a walking transect of the area and recorded the coordinates of water sources at 10m intervals on a GPS device. In Area 25 the primary sources were public water kiosks and wells (Figure 49a and figure 49b).

The water samples were tested for several heavy metals however, in this chapter I will focus on those of a priori interest: Lead [Pb], Cadmium [Cd], Arsenic [As], Nickel [Ni], Uranium [U], Chromium [Cr], Vanadium [V], Barium [Ba], Aluminium [Al], Fluoride [F^-] and hard water [$CaCO_3$]. Mercury [Hg] is a well-established nephrotoxin however, this could not be tested for as the laboratory apparatus was unavailable at the time of analysis. Furthermore, pesticides were not tested for as there were no reliable standards available at the time of analysis.



Figure 48. Photos of frequently used water sources in Bonje showing (a) a well; and (b) a borehole with an Afridev pump system



Figure 49. Photos of frequently used water sources in Area 25 showing (a) a water kiosk; and (b) a well

6.2.2.2 Study area water source profiles

Bonje lies on a plain, where the predominant soil has a near-neutral pH, and contains high calcium carbonate concentrations [307, 308]. Approximately 80km north of Bonje there is an extensive uranium deposit which is heavily mined and could affect the concentrations of heavy metals (specifically uranium and arsenic) in ground and surface waters[309]. Bonje is surrounded by a large rice paddy scheme which is approximately 1km by 2km in area (figure 50) where pesticides and herbicides are used seasonally, and agricultural run-off is directed to man-made moats (figure 50) outside the paddies which are near residential water sources.

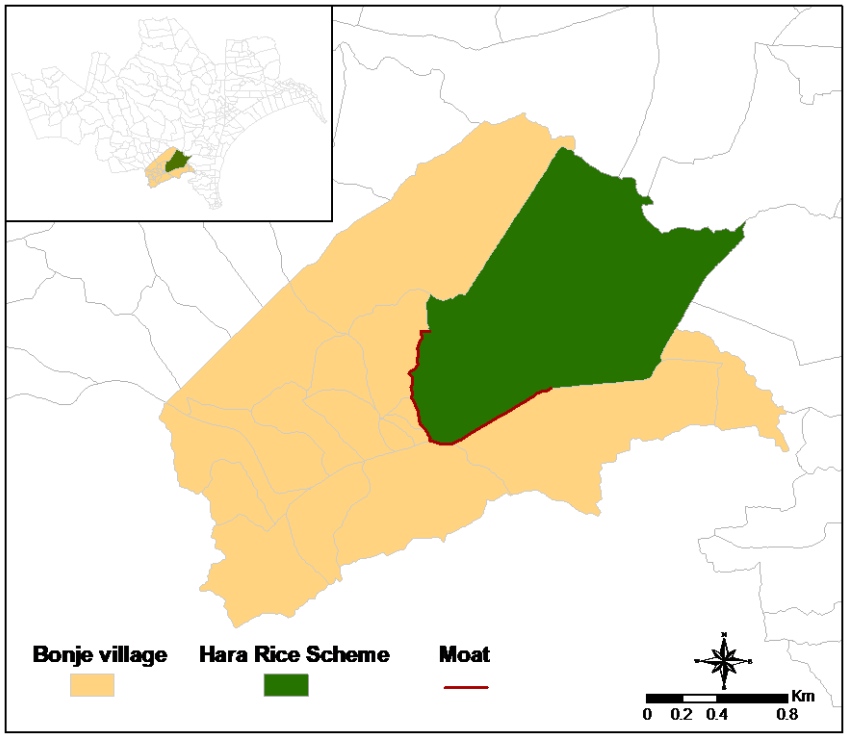


Figure 50. Map showing the location of Bonje village within the HDSS, and the location of Hara Rice Scheme

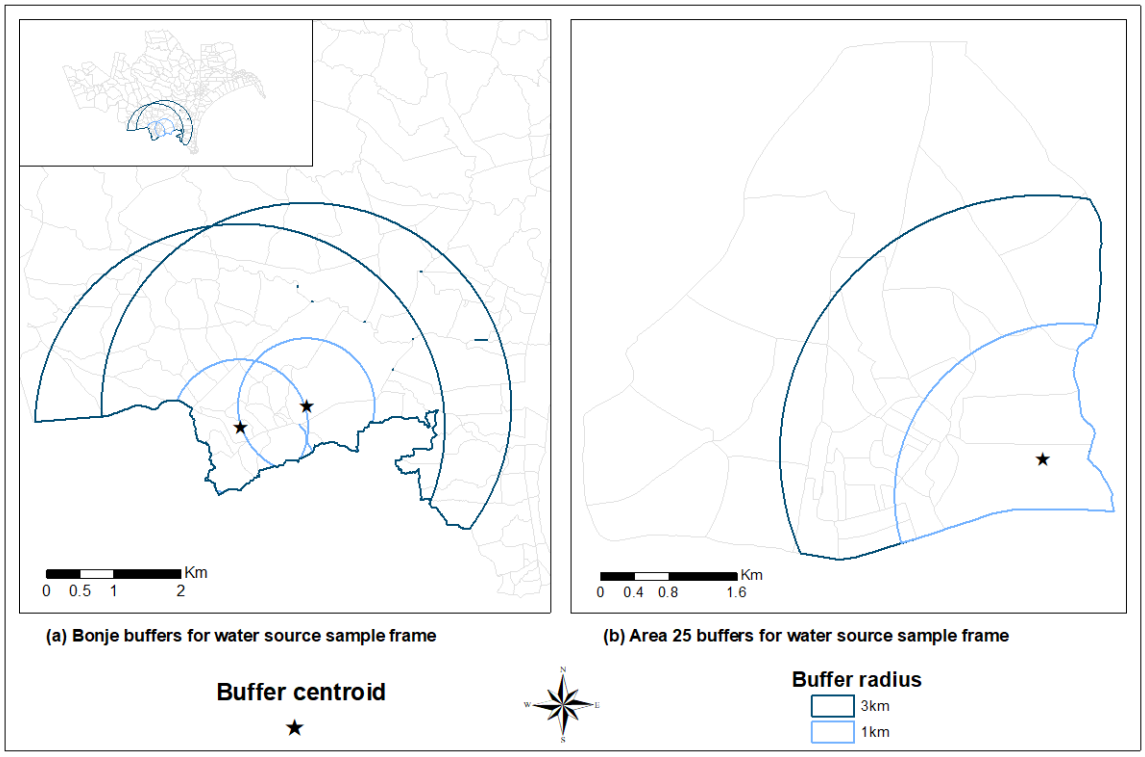
From my initial field research, the most commonly used water sources in Bonje are boreholes and wells. A borehole is a narrow shaft bored into the ground sub-surface, generally dug between 30-40 meters deep[310] and operated by a pump system. In Malawi, boreholes are increasingly being retrofitted with new Afridev pumps[310, 311] which increase the stroke flow, reducing the effort required to extract water. The pump head, handle and stand are made using various materials including stainless steel, fibreglass, polyvinylchloride or brass [311] (figure 48b). The wells across Bonje are approximately 12-20 metres deep, and a string and bucket system is used to extract water. The wells are classed as "covered" as the shaft is lined with brick to maintain the structure (figure 48a).

Area 25 lies on a plain where the dominant soil is an unconsolidated, weathered clay which has generally high Iron and Aluminium content [307, 312]. The Lilongwe Water Board supplies water to the city, which is extracted from the Lilongwe River and treated in the South East of the city. More recently however, access to this safe and reliable drinking water has decreased due to failing treatment facility conditions [313]. The Government of Malawi implemented the National Water Development programme which aimed to increase access to safe drinking water

through improved treatment plants and provision of water via communal kiosks. The kiosks are located mainly in the poorest areas of Lilongwe and are operated by a kiosk attendant providing water from a faucet in exchange for a fee (figure 49a). Conversely, wells in Lilongwe are dug approximately 12 metres deep and are located in resident's yards. As many people in Lilongwe cannot afford water kiosk rates, they tend to rely on these private hand-dug wells which are a mixture of covered (with a brick interior) and non-covered (non-lined) walls. Like Bonje, a bucket and string method is used to extract water (figure 49b).

6.2.2.3 Water source sampling strategy and data collection

For each study zone, I created a water source map using the mapping software ArcMap (version 10.5.1, ESRI, Redlands, US). To investigate water quality directly within and outside the participant study zones, I measured the central mean point of the participant housing clusters in Area 25 and Bonje and placed circular buffers of 1km and 3km radius around each mean point. In Bonje, participants were spread across a wider area than in Area 25, therefore I applied four buffers (2x 1km and 2x 3km) to capture the entire sample area (see figure 51 for buffers).



(a) Bonje buffers for water source sample frame

(b) Area 25 buffers for water source sample frame

Figure 51. Maps showing the sample frame buffers applied to each study zone for water collection in Bonje and Area 25

Water sources which fell within the 1km and 3km buffers served as the sample frame from which the random stratified sample was created. For each study zone, there were approximately 60 eligible water sources for selection. These were grouped by source ("borehole", "well", "kiosk") and assigned a number based on their location from West-East (in Bonje) and South-North (in Area 25) on the map. Using a random number generator in r studio, I created a random sample (n=20 per study zone, n=10 per buffer) stratified by water source.

Before sampling began, I conducted a reconnaissance visit around Bonje and Area 25 with field team members to assess road quality and accessibility, and to identify if any sources selected for sampling were defunct. This was necessary as there had been flooding in the period between sample design and collection. Unreachable and or/broken sources were removed from the sample frame, and the nearest functioning water source of the same type was selected for the final sample. Together with a water chemist from the Ministry of Agriculture Irrigation and Water Development, I collected samples in Bonje in March 2018, and in Area 25 in April. Water collection for each study zone took 3 days to complete. Figure 49 illustrates the final samples selected for Bonje and Area 25.

In the field I measured physicochemical parameters of each sample, including temperature, pH, dissolved oxygen, electrical conductivity, and turbidity as these denote overall water quality. These parameters have no known direct health impacts but can provide information on likely pollutants in the water. I extracted water from each source using the same methods as residents: pumping water directly from boreholes into containers, running water from the kiosk taps into the containers, and collecting water from wells using a bucket and string. Before each sample was collected, I washed the containers inside and outside with the potable water at each site, and I measured the physicochemical parameters of each sample using a handheld sonde. Parameters include temperature, pH, dissolved oxygen, electrical conductivity, and turbidity as these denote overall water quality (see appendix 31). To measure the physicochemical parameters, I collected a sample into a sterile 500ml polyethylene bottle, and the measurements were taken in duplicate and logged in a field tablet. The apparatus was thoroughly washed with spring water between each site, and these sample bottles were

labelled as “used” and stored in a separate box to be discarded at the laboratory. The physicochemical parameters measured, and apparatus used are detailed in table 42.

Table 42. Water source parameters measured in the field and apparatus used

Parameter	Device	Operating range
pH	Hanna HI9125 digital meter	-2.00-16.00pH
Temperature	Hanna HI-93510 handheld meter	-18.00^{o}C - 70^{o}C
Electrical conductivity	Hanna HI99300 digital meter	0-3999µS/cm
Turbidity	Hanna HI98703 field meter	0.00-1000 NTU (Nephelometric Turbidity Units)

I collected samples for heavy metals analysis in duplicate into sterile 500ml polyethylene bottles which after collection were sealed and stored in a polystyrene container in the field van. At the end of each day the bottles were stored at 5^{o}C in the local laboratory fridges (KPS in Bonje and MEIRU in Area 25).

When the collection was complete at each study zone, I prepared the samples for heavy metal analyses in the local laboratories ready for shipping to the UK, by preparing two test tubes of sample for each water source, both of which were filtered. I filtered bacterial cells from each sample using a single-use syringe (Discardit II Syringe 20ml Eccentric Disposable). To do this, I washed out each syringe with sample water three times, then filled it again and fitted a 0.22µm filter cap (FIL-S-PES-022-25-100 Syringe filter, PES, 25mm) on the end of the syringe. The water was filtered into a universal test tube (Sarstedt Tube Universal 25mm x 90mm 30ml Polypropylene Screw Cap).

After filtration, one tube was acidified with 1ml 5mol hydrochloric acid (HCL) to preserve trace metals in the solution and to prevent adsorption into the test tube walls in preparation for trace metals testing. All preparation was conducted wearing overalls and fresh powder-free gloves. Each universal tube was labelled with the site number on both sides on waterproof labels.

Samples were stored in a freezer at -24^{o}C for 3 days, with the screw cap on loosely to avoid bursting. Once they were ready to be shipped, I removed the samples from the freezer, then

sealed, and wrapped them in parafilm tape to stop leakage in transit. Each sample was double wrapped in labelled Ziplock plastic bags and placed into a polystyrene box packed with ice and transported to the University of Nottingham Environmental Sciences laboratory for trace metals and anion analyses. For the laboratory importation document and a DEFRA license see appendix 30 and 31.

6.2.2.4 Laboratory analyses

Samples were received by the laboratory in Nottingham, where the manager checked the universal tubes for any holes or leakages. Sample conditions were logged, and the tubes were stored in a freezer (-24^{o}C) for 2 weeks before analyses were conducted due to the busy laboratory timetable. Multi-element analysis of diluted solutions was undertaken by inductively coupled plasma mass spectrometry (ICP-MS), and Ion chromatography (IC) analyses were conducted. For details of these methods and the standards used, see appendix 32-34. The laboratory team logged results in an online database and sent to me via a .csv file.

6.2.2.5 Water source assignment

The aim of water source collection was twofold – to measure heavy metals concentrations, and to trace each participant to a water source to assign them to a heavy metal concentration. In order to assign a water source (and heavy metals value) to each participant, I created buffers around each water source in the study zone and assigned participants to the closest source. The buffer size was defined as the average distance a resident would walk to collect water in each study zone, both Bonje and Area 25 which was estimated by field workers who live in and around these zones. In Bonje the average distance was 500m, whereas in Area 25 it was 200m. Using the spatial join tool in ArcGIS, participants were assigned to their nearest source.

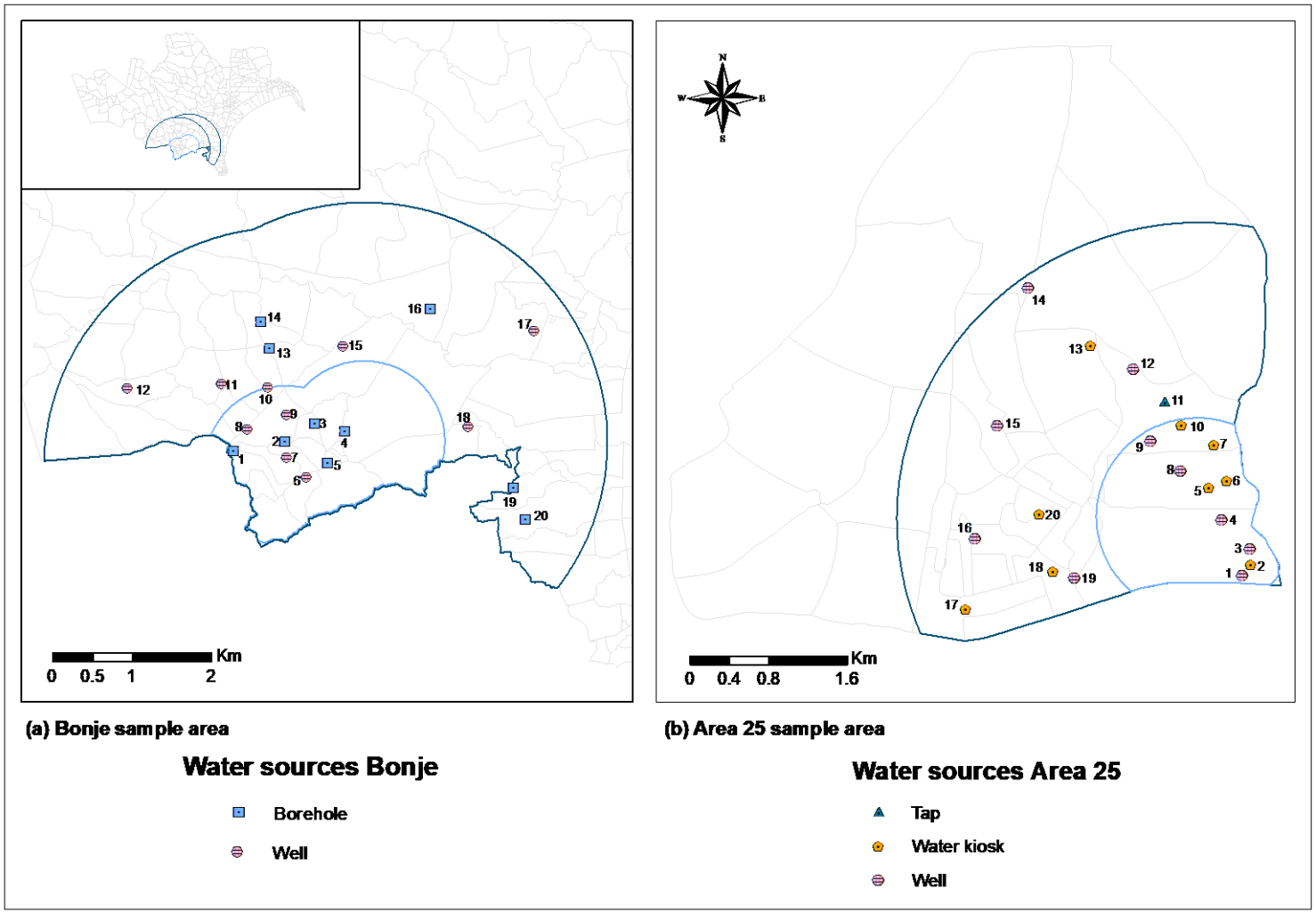


Figure 52. Water sources sampled in (a) Bonje and (b) Area 25. Sources are numbered according to the order in which they were sampled

6.2.3 Environmental study part 3: Satellite imagery assignment

I assigned satellite-derived environmental data (listed in table 43) to each participant across Bonje and Area 25 in order to conduct regression analyses to investigate possible associations between the environment and eGFR. I used the same data processing and participant assignment methods as described in chapter 4. For the heat index models I log transformed the data due to skew (see appendix 35) and used the same techniques as in chapter 4 to model a surface. However, I used a different land cover dataset for Malawi which I will describe below.

6.2.3.1 Land cover data

The ESA-S2-LC20 2016 land cover prototype for Africa describes the land surface at 20m spatial resolution[314], and was produced at a regional level with 22 different classes at first, that were simplified to 10 generic classes. The S2 prototype was generated from one year of Sentinel-2A observations from December 2015 to December 2016 using random forest Machine Learning (ML) algorithms.

Table 43. Summary of satellite-derived environmental data sources used in environmental analyses for Bonje and Area 25, Malawi

Risk factor	Data source	Spatial resolution	Time	Coverage	Processing method
Meteorology (temperature, humidity)	ECMWF ERA-5	31 x 31 km 2m vertical accuracy	1979-present	Global	Empirical Bayesian Kriging
Altitude	Shuttle Radar Topography Mission	30 x 30 m 20m vertical accuracy	2000	Global	Mosaic
Land cover	ESA-CCI LC – S2 Prototype Land Cover 20m Map of Africa	20 x 20 m	2016	Africa	Extract by mask

To assign the modal land cover classes to participants, I applied a 20m circular buffer around each participant coordinate in ArcMap in order to match the resolution of this satellite dataset.

In the following section I describe how the data sources discussed above were coded, re-grouped and merged into a harmonized dataset.

6.2.3.2 Coding, categorizing, and merging data

After questionnaire collection was complete, the documents were scanned into the shared database. I reviewed each, checking for missing data or illegible free text answers, and coded the variables. In order to create more parsimonious models, I re-grouped some larger variables with multiple sub-categories into broader variables. Occupation status was re-grouped into discrete categories “agricultural” (including home makers who conduct intensive subsistence farming), “non-agricultural”, “vendors” (shop and street vendors and those carrying items or people), and “non-workers - able” (encompassing students). Water sources were grouped into “borehole” (which included piped in–o home/yard), “well” (encompassing all naturally sourced and non-covered water sources such as lakes and rivers), and “kiosk” (which included bottled water). Alcohol and smoking variables were separated in these analyses as smoking is a major confounder for Cadmium in blood and needed to be adjusted for separately. As in chapter 4, land cover categories were re-grouped into “urban” and “cropland”. Altitude and heat index were modelled as continuous variables. Due to skewness of the blood and water heavy metal data, all values were log transformed and scaled.

Following this, I merged the satellite and measured environmental variables (assigned water metals, altitude, heat index and land cover) with the environmental questionnaire data using the STID as the common identifier to form a harmonised dataset.

6.2.3.3 Model building and statistical methods

As reported in chapter 5, results indicate that overall CKD/u is not present in the Malawian population (at least to the extent to which it was investigated), which, coupled with a small sample size caused concern over sparse data. I therefore conducted variable profiling of all covariates in the harmonized environmental questionnaire with linear and logistic regression models which were minimally adjusted for age, sex, and study zone. Those which had an independent association with eGFR were then mutually adjusted in a second model. I tested all models for multicollinearity, using the variance inflation factor (VIF) >10 as the threshold for exclusion.

Sensitivity analyses

As in chapter 4, I assessed the gains from using the Empirical Bayesian Kriging method, I conducted a sensitivity analysis by comparing the original heat index data with more simplistic interpolation methods inverse distance weighting, bilinear interpolation, and kriging. I selected a smaller search radius of 4 neighbours and restricted the data points to altitudes within 100m of those in the target study zones across Malawi.

6.2.3.4 Water sources

All heavy metal concentrations were compared with the World Health Organization (WHO) and Malawi Bureau of Standards (MBS) permissible limits [127, 315]. A Kruskal-Wallis test was used to differentiate heavy metals concentrations between Bonje and Area 25; and then between water sources, based on median values. The null hypothesis (H_o) was that there was no difference in median values between study zones. Significant differences were determined at the $p < 0.05$ 95% confidence level. To avoid type I error, I conducted a post-hoc Bonferroni correction to the data. I applied Principal component analysis (PCA) on the combined heavy metals dataset (Bonje and Lilongwe) to identify metals and/or interactions of metals which represent the largest variation of heavy metal concentration in the water sources. PCA is a technique used to identify dominant patterns in multi-dimensional data which reduces the dimensions of a dataset, whilst retaining the information it contains by finding the weighted sums of variables that capture as much of the variance in the data as possible [316]. I selected the optimal coordinates using a scree plot to determine the number of components to retain. Varimax rotation of the retained standardized component loadings was conducted for maximizing the variation among the variables under each factor, and for ease of interpretation I rescaled the scores of each principal component to lie in the range of 1-100. To assess the overall difference in heavy metals concentrations between Bonje and Area 25, I applied a t-test to each PCA output score, using an “urban” “rural” identifier in the dataset. To reduce the number of tests and avoid a type I error I did not apply the t-test to each metal separately.

6.2.3.5 Biomarkers

As blood reference values for heavy metals are difficult to quantify, I used a range of data sources including World Health Organization and Centre for Diseases and Control guidelines, to decipher a mean reference value for each metal.

I applied Bayesian Kernel Machine Regression (BKMR), which estimates the multivariable exposure-response function of the blood heavy metal biomarkers with the outcome eGFR, allowing for potential and synergistic effects amongst mixture components. The model was estimated by Markov chain Monte Carlo (MCMC) algorithm[317]. In this approach, the eGFR is regressed on a flexible function of the mixture using a kernel function[318]. This allows for estimation of nonlinear and non-additive dose-response functions for a set of correlated exposures, and simultaneously provides measures of variable importance (posterior inclusion probabilities). For Bonje and Area 25, I modelled blood heavy metals as natural log-transformed and centred continuous variables. This transformation creates a common scale, accounting for the left skewedness of the metal's concentrations. eGFR values were relatively normally distributed and modelled as scaled continuous variables. Water was not included in the mixtures because multiple participants were assigned to a water source (and thus heavy metal variable), resulting in multiple repetitive values which could not be modelled efficiently. The following covariates were included in all models to adjust for confounding: sex, age, and smoking status (yes vs. no). There were no missing covariates in this dataset.

To begin, I conducted multivariable linear regression analysis to investigate associations between eGFR (as a continuous variable), heavy metals in the blood, and the confounders listed above. I then applied the BKMR model to the variables as shown below:

Equation 8. BKMR formula

$$\mathrm{Yi} = \mathrm{h}\,(Z_{i1}, \ldots, Z_{iM}) + \beta^{Y} Z_i + \epsilon i \qquad (318)$$

Where the function h() is an exposure-response function accommodating for nonlinearity or interaction amongst mixture components (z_{i1},…,z_{iM}), deciphered using a kernel exposure-response machine representation in a linear mixed model. The resulting models assume that two subjects with similar exposures (e.g. z_{i1} is "close" to z_{i2}) will have similar estimated health effects (e.g., $h_1 = h(z_1)$ "close" to $h_2 = h(z_2)$ for individuals *1* and *2*). The similarity between individuals' exposure profiles is measured using the kernel function[318].

To fit the BMKR model, I used the Gaussian kernel function to process the high dimensional parameter space in the exposure-response function[317, 319]; this function was used to account for the possibly complex relationship that the metal exposure variables have on eGFR. The

Gaussian kernel has also been shown to perform well in simulated exposure-response scenarios[318].

6.2.3.6 Correlations between water and blood heavy metals

I conducted a Pearson Correlation Coefficient to determine potential correlations between blood and water heavy metals which were logged and scaled due to the different measurement scales for each matrix. The covariance is a measure of this relationship and depends on the variability of each of the two matrices. Finally, to assess if blood and water metals are associated with eGFR, I modelled these matrixes together, adjusting for potential confounders age, sex, and smoking. In these analyses, all metals were included except for hardness, as this cannot be measured in bio samples.

All statistical analyses were conducted in r Studio version 3.5.1.

6.3 Results

6.3.1 Study population characteristics

The nested study population comprised 367 adult participants over 18 years (males =146; females =221). This sample size was smaller than originally anticipated as the DEGREE study was terminated early in June 2018 due to funding shortages.

Study participants were not excluded if they were diabetic, hypertensive and/or proteinuric as the findings in chapter 5 denote that CKD is likely not a health burden in this population. This study was therefore a cross-section of a population sub-sample. Participants missing serum creatinine measures were excluded however, as this were required to calculate eGFR. Figure 53 is a study flowchart showing the final sample size after exclusions.

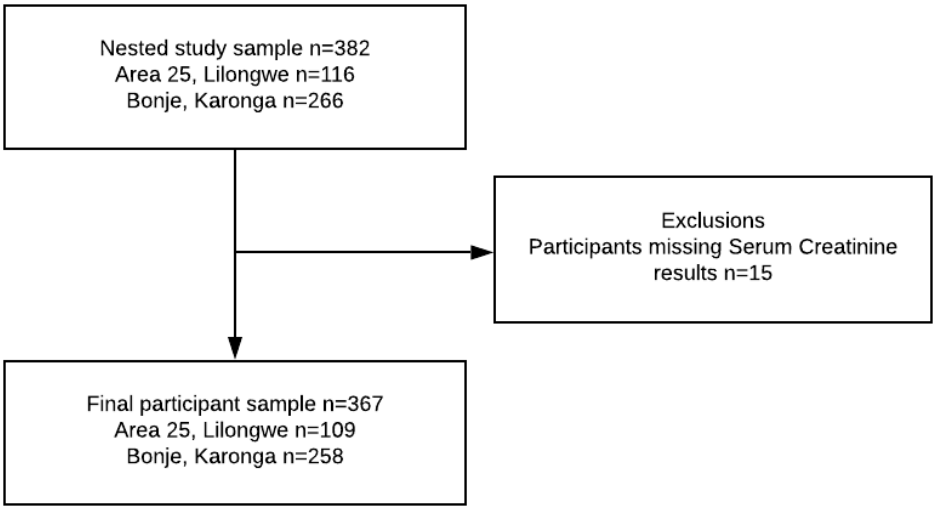


Figure 53. Study flowchart with exclusion criteria for the Malawi nested study population sample (n=367)

6.3.2 Socioeconomic and environmental characteristics

The mean (standard deviation (±SD)) participant age was 36.0±4.4 years. Mean BMI was 24.0±4.4 kg/m^2, and mean FFM was 50.5±71.8 kg/m^2. Mean systolic and diastolic blood pressure was 122.1±17.2 mm Hg, and 73.9±11.6 mm Hg respectively. Mean fasting plasma glucose was 4.6±0.69 mmol/l and the median (interquartile range (IQR) albumin/creatinine ratio (ACR) was 0.8 (0.3, 1.3) mg/mmol. In this nested population 3 participants (0.8%) were diabetic, 53 (14%) were hypertensive, and 3 were proteinuric (0.8%).

Approximately 65% of participants were conducting daily agricultural activities, either in farm plots or home garden plots (females n= 149 [67%], males n= 91 [62%], and the majority of the population were in year-round employment (n=294 [80%]) as opposed to seasonal employment. Overall, 50% of the participants had completed >10 years of formal education, and more than half of the participants were in the higher monthly household income category (MK>20,000) (see table 33).

6.3.3 Environmental variables

Figure 54 (a-c) is a map showing satellite-derived environmental maps of Malawi. For the heat index kriging model fit for Malawi please see appendix 36. Bonje is situated at a relatively low

altitude ranging between 500-531 metres (figure 55c). This population sample is entirely rural, with all participants living within 20m of "grassland" and "cropland" (figure 55a).
As expected with a small study area (2km), there was a very small heat index range of 21.7-22.2^{o}C across the study zone (figure 55b) with an increasing gradient from the higher altitude west to the lake-neighbouring east of the zone (see figure 55a). Bilinear heat index results ranged from 22.5- 23.6 ^{o}C. Comparatively, Area 25 is situated at a higher altitude, between 1099-1125m (figure 56c), and has a lower (and virtually no major gradient) heat index of 19.3-19.4^{o}C (figure 56b). Bilinear results ranged from 19.4-19.5 ^{o}C.

Table 44 shows the model fit for the hold out validation summary statistics for the Malawian heat index model which show that there was an overall good fit (R^2= 0.951; RMSE= 0.141) of the interpolated heat index data. See appendix 35 for cross validation model fit regression line.

Table 44. Empirical Bayesian Kriging hold out validation summary for heat index across Malawi

Prediction errors empirical Bayesian kriging	Malawi
Mean absolute error	0.121
Root Mean Square error	0.141
R^2	0.951

The predicted standard error map (figure 57) across Malawi follows a similar pattern to those seen in Sri Lanka and India in chapter 4. Overall, the highest uncertainty was observed in the highest altitude regions of the country in the North (1569-2304m) where the standard error ranged from 0.901^{o}C -0.955^{o}C. Across the study zones the uncertainty was marginally higher in Bonje (0.541 ^{o}C – 0.695 ^{o}C), than that observed in Area 25 (0.520 ^{o}C – 0.523 ^{o}C).

In Area 25 the land cover categories are mixed, with 81% (n=88) of the population living in the vicinity of "urban" land cover, and 19% (n=21) living close to "cropland" (figure 56a, and appendix 34 for model fit), and water kiosks were the most used water sources with 65% (n=71) (see table 47). In Bonje most of the study population reported using boreholes as their primary water source (84%, n=218). Across both study sites large majority of the population

(84%, n=307) did not report using any water sanitation treatments, but amongst those who did, WaterGuard - a local chlorine product promoted by Population Services International (PSI) [320] - was the most common treatment (n=51 [14%]).

Of the nested study participants, approximately 58% (n=214) answered the agricultural questionnaire section, of which 98% (n=210) were Bonje residents. Half of the farming practices reported were for subsistence purposes (49.5%), followed by commercial farming (29.9%). The most cultivated crops were rice, maize, groundnuts and cassava, and few participants reported self-application of pesticides (n=22 [10.2%]). Very few of the 22 (10.2%) participants reported using PPE when applying pesticides, which include facemasks (n=3), overalls (n=4) and footwear (n=3).

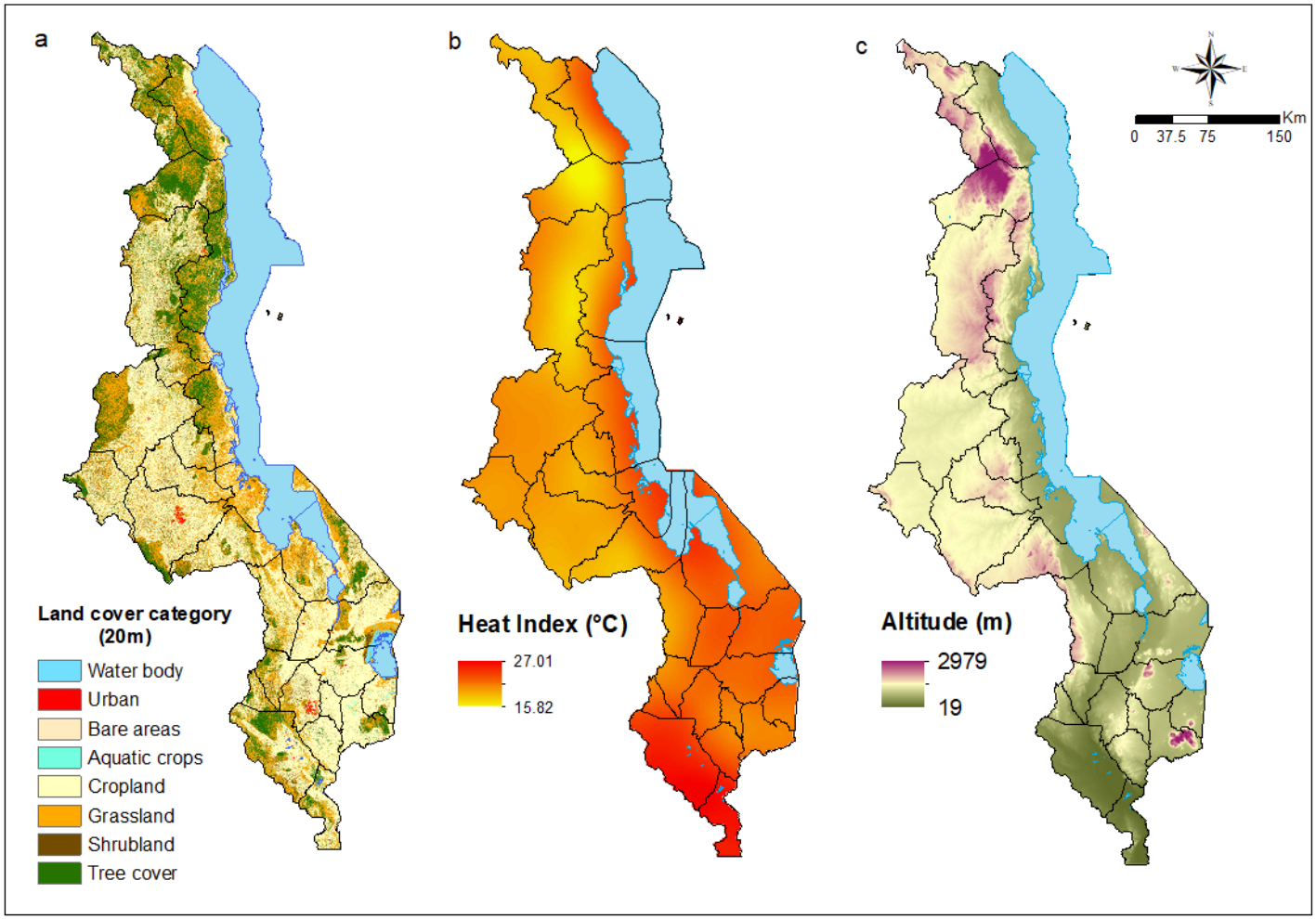


Figure 54. Land cover (a), heat index (b) and altitude (c) surfaces of Malawi modelled using Empirical Bayesian Kriging; refer to figure 40 for study zone locations

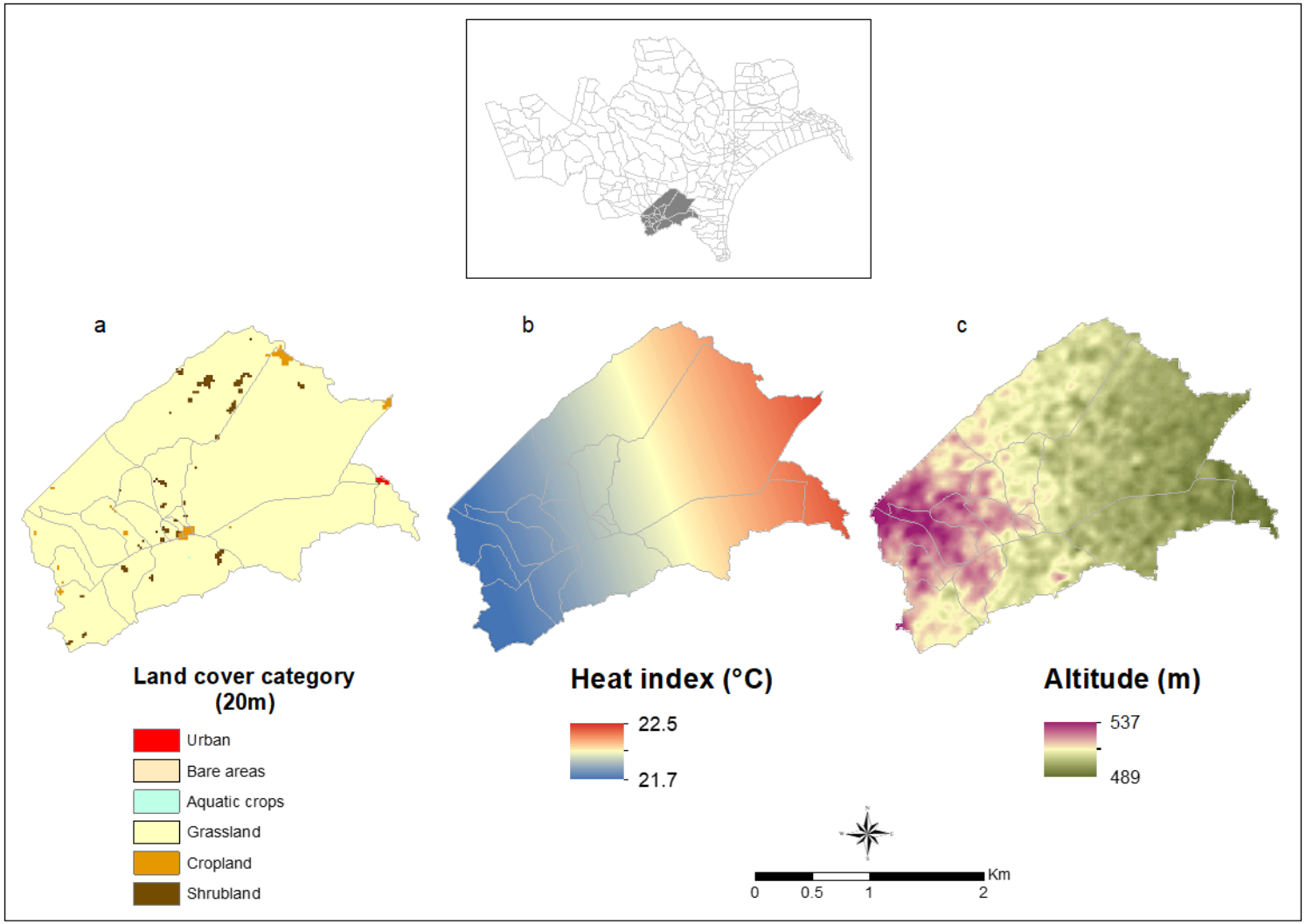


Figure 55. Land cover (a), heat index (b) and altitude (c) surfaces of Bonje modelled using Empirical Bayesian Kriging

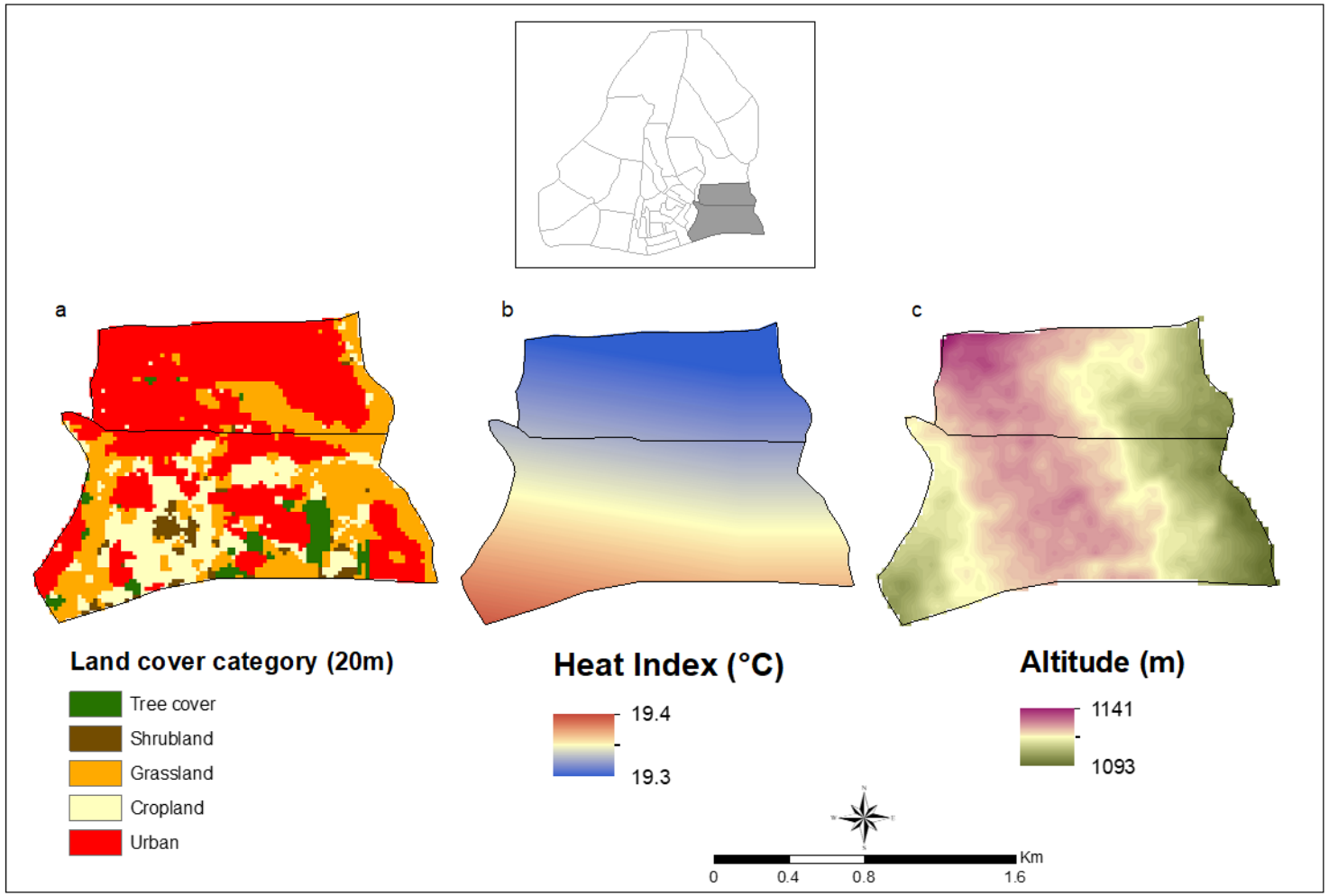


Figure 56. Land cover (a), heat index (b) and altitude (c) surfaces of Area 25 modelled using Empirical Bayesian Kriging

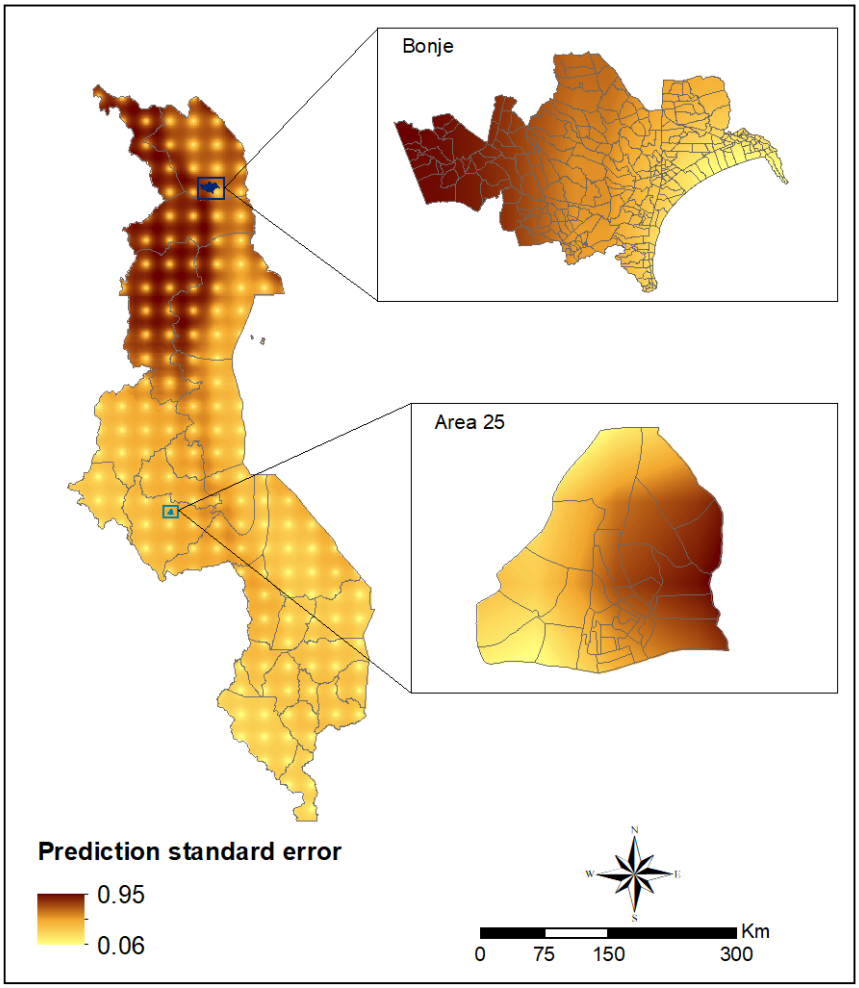


Figure 57. Prediction standard error maps for Empirical Bayesian Kriging interpolation of heat index data across Malawi

Sensitivity analysis

As in chapter 4 I conducted a sensitivity analysis in order to understand whether the EBK interpolation method (scaled down from the national to study zone level) was "fit for purpose" as a proposed interpolation method for the ERA-5 dataset. For data distribution of the heat index data restricted to altitudes within 100m of the study zones, and model fit for the interpolation methods used in this sensitivity analysis please see appendix 37.

Table 45 shows the comparative results from the validation statistics show that RMSE and MAE values were the lowest for the bilinear interpolation model, (0.005 and 0.003 respectively), with the highest values (and therefore error) seen in the kriging model (see table 45). Both IDW and bilinear interpolation had the highest R^2 values (0.999 respectively) however the lower error values for bilinear interpolation would suggest this method provides the pest prediction of the heat index surface for Malawi.

Table 45. K-fold cross validation summary of the performance of interpolation methods to predict heat index across Malawi

Interpolation method	Efficiency	Error	
	R^2	RMSE	MAE
IDW	0.999	0.009	0.006
Kriging	0.988	0.011	0.004
Bilinear interpolation	0.999	0.005	0.003
EBK	0.997	0.012	0.008

R^2 = coefficient of determination, RMSE = root mean square error, ME = mean error, MRE = mean relative error, IDW: inverse distance weighting; EBK- empirical Bayesian kriging

In terms of Pearson correlation (table 46), the results of, Kriging, IDW, bilinear interpolation are significantly correlated with the original ERA-5 modelled data at 0.01 level (two-tailed). The correlation coefficient between bilinear interpolation and IDW interpolation and the original ERA-5 modelled data were the highest at 0.99 respectively.

Table 46. Correlation analysis of interpolation methods across Malawi

Malawi-wide	Original HI data	Kriging	Bilinear	IDW	EBK 10-15 neighbours
Pearson correlation	1.00	0.98	0.99	0.99	0.99
p	0.000	0.001	0.000	0.000	0.000

Like the model fit results seen in chapter 4 for India and Sri Lanka, overall, among the interpolation methods, the performance of bilinear interpolation was best in comparison with other interpolation models.

6.3.4 Mean eGFR and prevalence of reduced eGFR

The mean eGFR was 117.0±16.5 mL/min/1.73m^2 (figure 58 and table 47), and there was an inverse relationship observed between eGFR and increasing age and BMI. Agricultural workers, specifically those cultivating maize, and those applying pesticides without personal protective equipment had a low eGFR. Seasonal labourers also had a lower eGFR than all-year workers. Participants reporting employment related symptoms such as raised heart rate, thirst and muscle cramps also had a lower eGFR than those who did not. For water sources, the lower eGFR values were observed amongst participants using water kiosks in Area 25, and wells in both study zones, and little difference was observed between eGFR and daily water intake. Furthermore, participants earning a higher monthly income (MK>20,000) had a lower eGFR, which could be indicative of having access to more unhealthy lifestyle choices such as a high fat diet. Site-specifically, the mean eGFR was lower amongst urban participants (114.5(±14.3)) than rural participants (118.1(±12.2)), and no difference in mean eGFR was observed between males and females. Contrary to the literature, participants residing at higher temperatures and lower altitude had a higher eGFR.

Across this sub-sample, the number of participants with an eGFR<90 was small (n=23). The higher prevalence of eGFR<90 was observed in Bonje (6.5%), and sex-specifically the prevalence was the same between males and females.

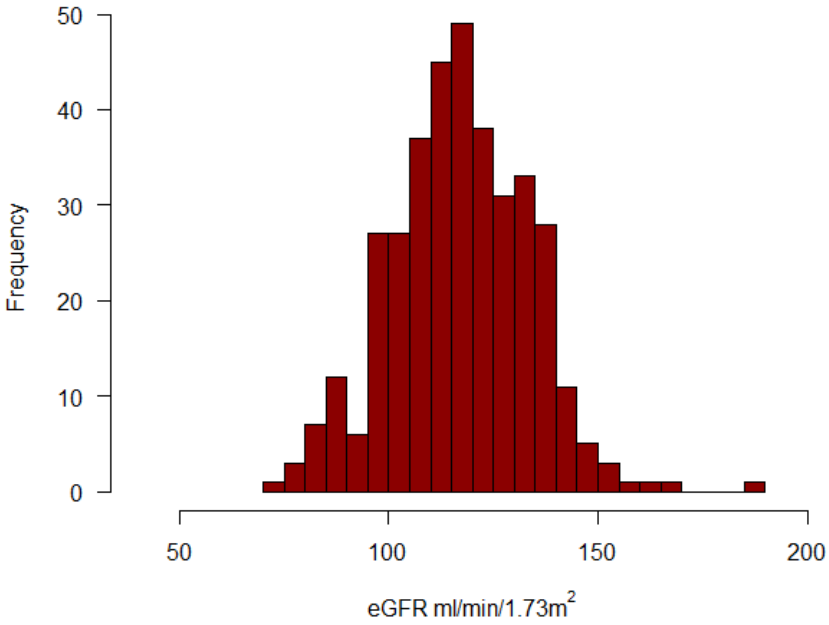


Figure 58. Histogram of eGFR distribution in the nested sample population in Malawi (n=367)

Table 47. Socio-economic and demographic characteristics of study participants in the Malawian sub-sample (n=367)

Variable		eGFR n= 367	eGFR categories, n(%)[b] n= 23	n= 344
	n(%)[a]	Mean (SD)	<90	>=90
Age (years)				
<20	24 (7)	134.2 (15.4)	0	24 (100)
20-29	111 (30)	128.6 (11.4)	0	111 (100)
30-39	98 (27)	117.0 (10.8)	0	98 (100)
40-49	68 (19)	112.1 (10.7)	0	68 (100)
50-59	27 (7)	103.2 (9.6)	3 (11)	24 (89)
60+	39 (11)	91.5 (9.6)	20 (51)	19 (49)
Sex				
Female	221 (60)	117.0 (16.0)	11 (5)	210 (95)
Male	146 (39)	117.1 (17.2)	12 (8)	134 (92)
Study site				
Urban (Area 25)	109 (29)	114.5 (14.3)	6 (6)	103 (94)
Rural (Bonje)	258 (70)	118.1 (17.2)	17 (7)	241 (93)
Education (years)				
≤5	36 (10)	106.8 (17.7)	7 (19)	29 (81)
>5≤10	155 (42)	117.2 (14.8)	7 (5)	148 (95)
>10	176 (48)	118.9 (16.9)	9 (5)	167 (95)
Occupation				
Not working (able)	37 (10)	129.5 (16.3)	0	37 (100)
Agricultural worker	240 (65)	116.2 (16.4)	17 (5)	228 (95)
Non-agricultural worker	43 (11)	118.7 (13.5)	4 (9)	39 (91)
Street/shop vendor	42 (11)	116.1 (15.4)	2 (5)	40 (95)
Work pattern				
All year	294 (80)	116.4 (15.8)	18 (5)	278 (95)
Seasonal	38 (10)	113.7 (18.8)	5 (13)	33 (87)
Missing data	35 (9)	-	-	-
Rest frequency				
Never	95 (25)	116.2 (16.4)	6 (6)	89 (94)
Hourly	119 (32)	120.5 (17.8)	6 (5)	113 (95)
Every 2+ hours	153 (41)	114.9 (14.8)	11 (7)	142 (93)
Sweat				
Yes	247 (67)	115.9 (16.5)	20 (8)	227 (92)
No	120 (32)	199.2 (16.2)	3 (2)	117 (98)
Raised heart rate				
Yes	97 (26)	115.5 (15.6)	7 (7)	90 (93)
No	270 (73)	117.6 (16.7)	16 (6)	254 (94)
Thirst				
Yes	267 (72)	116.1 (15.9)	18 (7)	249 (93)
No	100 (27)	119.2 (17.6)	5 (5)	95 (95)
Muscle cramps				
Yes	150 (40)	113.7 (15.2)	11 (7)	139 (93)
No	217 (59)	119.3 (16.9)	12 (6)	205 (94)
Household monthly income (MK)[c]				
Unknown	7 (2)	116.9 (11.5)	0	7 (100)
MK 0 ≤20,000	168 (45)	119.1 (18.2)	10 (6)	158 (94)
MK >20,000	192 (52)	115.2 (14.8)	13 (7)	179 (93)
Healthy lifestyle choices				

Non-smoker/ never drink alcohol	279 (76)	116.7 16.4)	18 (6)	261 (94)
Smoker/ alcohol drinker	88 (23)	117.9 (16.4)	5 (6)	83 (94)
Body mass index (kg/m^2)				
Underweight (≤18.5)	15 (4)	119.0 (22.8)	1 (7)	14 (93)
Normal (>18.5 - ≤ 25)	234 (63)	119.1 (16.4)	10 (4)	224 (96)
Overweight (>25 - ≤30)	80 (21)	111.5 (13.9)	9 (11)	71 (89)
Obese (>30)	38 (10)	114.8 (16.7)	3 (8)	35 (92)
Fat-free mass (kg)				
1st tertile (≤37)	61 (33)	112.6 (17.6)	6 (10)	55 (90)
2nd tertile (>37 - <45)	133 (96)	116.3 (16.9)	10 (8)	123 (92)
3rd tertile (≥45)	173 (93)	119.1 (15.3)	7 (4)	166 (96)
HIV status				
Positive	0	-	-	-
Negative	276 (75)	116.5 (15.9)	17 (6)	259 (94)
Unknown	91 (24)	118.7 (17.8)	6 (7)	85 (93)
Primary water source				
Well	46 (12)	113.5 (15.5)	4 (9)	42 (91)
Kiosk	71 (19)	112.5 (14.7)	6 (8)	65 (92)
Borehole	250 (68)	118.9 (16.8)	13 (5)	237 (95)
Water intake				
<5 cups	262 (71)	117.1 (16.8)	17 (1)	259 (99)
>5 cups	81 (2)	116.1 (14.7)	6 (7)	75 (93)
Do not know	24 (6)	119.0 (18.1)	1 (4)	23 (96)
Water treatment				
Boiling	9 (2)	110.7 (14.2)	2 (22)	7 (78)
WaterGuard	51 (13)	117.5 (13.5)	2 (4)	49 (96)
No treatment	307 (83)	117.1 (16.8)	19 (6)	288 (94)
Weekly fish intake				
None	13 (4)	113.6 (15.4)	1 (8)	12 (92)
1-4 times	246 (67)	117.4 (16.3)	11 (4)	235 (96)
>4 times	108 (29)	115.8 (16.8)	11 (10)	97 (90)
Weekly meat intake				
None	59 (16)	113.8 (14.4)	3 (5)	56 (95)
1-4 times	284 (77)	117.5 (16.9)	20 (7)	264 (93)
>4 times	24 (7)	119.6 (14.8)	0	24 (100)
Land cover category				
Urban	88 (23)	113.5 (14.2)	5 (6)	83 (94)
Cropland	279 (76)	118.1 (16.9)	18 (6)	261 (94)
Heat index				
<21	109 (29)	114.5 (14.3)	6 (5)	103 (95)
>21	258 (70)	118.1 (17.2)	17 (7)	241 (93)
Altitude				
<1000m	258 (70)	118.1 (15.9)	17 (7)	241 (93)
>1000m	109 (29)	114.5 (17.8)	6 (6)	103 (94)
Farming category				
Commercial	64 (17)	118.7 (14.8)	2 (3)	62 (97)
Small holder	44 (11)	119.9 (18.2)	3 (7)	41 (93)
Subsistence	106 (28)	115.9 (17.1)	9 (8)	97 (92)
Does not farm	153 (41)	-	-	-
Rice cultivation				
Yes	200 (54)	117.5 (16.7)	14 (7)	186 (93)
No	14 (3)	119.6 (17.4)	0	14 (100)
Maize cultivation				

Yes	174 (47)	111.9 (11.1)	11 (6)	163 (94)
No	40 (11)	117.6 (16.7)	3 (7)	37 (93)
Cassava cultivation				
Yes	175 (47)	118.1 (16.5)	12 (7)	163 (93)
No	39 (10)	115.6 (17.8)	2 (5)	37 (95)
Groundnut cultivation				
Yes	56 (15)	122.3 (14.9)	4 (7)	52 (93)
No	158 (43)	115.9 (17.0)	10 (6)	148 (94)
Pesticide self-application				
Yes	22 (5.9)	115.4 (18.5)	3 (14)	19 (86)
No	192 (52)	117.8 (16.5)	11 (6)	181 (94)
PPE: facemask				
Yes	5 (1)	116.8 (28.2)	3 (60)	2 (40)
No	17 (4)	114.9 (16.6)	1 (6)	16 (94)
PPE: overalls				
Yes	4 (1)	111.3 (26.8)	2 (50)	2 (50)
No	18 (5)	111.2 (17.02)	1 (6)	17 (94)

[a] percentage in columns; [b] percentage in rows; [c] Exchange rate (MK to USD) 0.001 at time of questionnaire

6.3.5 BKMR analyses: blood

The following section reports heavy metals results from full blood samples in Bonje and Area 25 (table 48), and details linear regression and BKMR analysis which was applied to this data. Overall, across each study zone the mean heavy metal concentrations did not exceed permissible limits; however, concentrations were generally higher in Bonje. Furthermore, In both zones there were few participants with blood concentrations which exceeded limits, specifically for Chromium (n=34), Nickel (n=3), Aluminium (n=3) and Barium (n=2) in Bonje; and Vanadium (n=22) and Barium (n=1) in Area 25.

Table 48. Descriptive blood metals results in Bonje and Area 25

Metal	N	Mean	Min	Max	Median	SD	Reference values* (µg/l)	Exceedance of reference values (n)
				Bonje				
As	258	1.74	0.24	2.79	1.70	0.33	10.0	0
Cd	258	0.16	0.001	0.67	0.10	0.12	50.0	0
Pb	258	16.65	2.58	104.12	14.30	11.66	210.0	0
Ba	258	3.28	1.95	12.16	3.10	0.96	1.2-8.7	2
U	258	0.02	0.001	0.13	0.00	0.01	<0.5	0

V	258	0.06	0.001	0.57	0.57	0.09	0.05-0.08	0
Al	258	14.91	0.001	65.38	16.20	11.42	10.0-50.0	3
Cr	258	1.19	<0.001	9.05	0.50	1.73	2.0-3.0	34
Ni	258	1.43	0.78	8.29	1.30	0.58	1.0-3.0	3
				Area 25				
As	109	1.59	0.90	2.88	1.60	0.39	10.0	0
Cd	109	0.08	0.001	0.58	0.10	0.07	50.0	0
Pb	109	14.88	3.09	67.94	11.30	11.05	210.0	0
Ba	109	3.02	1.84	12.77	2.90	1.11	1.2-8.7	1
U	109	0.004	0.001	0.02	0.00	0.004	<0.5	0
V	109	0.08	0.001	5.22	0.00	0.50	0.05-0.08	22
Al	109	6.44	0.001	41.02	0.00	10.41	10.0-50.0	0
Cr	109	0.15	<0.001	1.01	0.00	0.28	2.0-3.0	0
Ni	109	1.39	0.58	2.35	1.40	0.35	1.0-3.0	0

To investigate associations between blood metals and eGFR, I conducted multivariable linear regression models stratified by study zone (table 49). Among the potential confounders, age was the most strongly associated with eGFR in both study zones; however, no associations were observed between any heavy metal and eGFR.

Table 49. Multivariable liner regression model of eGFR and heavy metals blood concentrations stratified by study zone in Malawi

Variable	Bonje		Area 25	
	β[a]	95% CI	β[a]	95% CI
Age (10-year increase)	-9.05***	-9.90, -8.20	-9.25***	-10.89, -7.61
Sex				
Male	-0.33	-3.00, 2.32	-0.50	-4.68, 3.68
Female	REF	REF	REF	REF
Smoker				
Yes	1.82	-2.92, 6.57	-3.2123	-23.11, 16.68
No	REF	REF	REF	REF
Blood metals (Log_e)				
Al	-0.92	-2.75, 0.90	-1.00	-4.32, 2.30
V	0.71	-0.75, 2.17	0.41	-1.73, 2.56
Pb	-0.12	-1.44, 1.19	-1.43	-3.37, 0.49
Cd	1.08	-0.51, 2.69	-0.25	-1.89, 1.38
Ni	-0.40	-1.72, 0.91	-0.14	-2.02, 1.73
Ba	-0.25	-1.63, 1.12	1.59	-0.16, 3.35
U	1.60	-1.48, 4.68	4.35	-1.26, 9.97
Cr	0.52	-1.90, 2.96	-4.77	-10.74, 1.18
As	-0.61	-1.91, 0.69	1.37	-0.60, 3.34

I applied BKMR models to estimate the exposure-response function of the nine heavy metals on eGFR, relaxing the assumptions of linearity and interactions of linear regression. I firstly

explored the cumulative effect of blood metal mixtures in eGFR, with concurrent changes in each mixture component from their median value. I then investigated the effect of a change in each metal by an inter quartile range (IQR) on eGFR.

Following this, I modelled the univariate dose-response relationships of mixture components on eGFR, and finally, potential interactions of metals; models which both estimate the predicted eGFR for each metal quartile, whilst setting the other metals at the 25^{th}, 50^{th} or 70^{th} percentile. Inferences on mixture effects were obtained by computing posterior mean estimates (and 95% posterior credibility intervals)[321] of the change in eGFR associated with changes in each level of each mixture component.

Figure 59 shows a summary of the overall association of the heavy metal mixture in Bonje which is defined by the change in the scaled eGFR value which is associated with a change of each metal from the 25^{th} to 75^{th} percentile, compared to when the metals are held at their median. There is no significant effect of the metals on the eGFR, however there is a weak negative trend which was investigated to see which metals might be driving this. Figure 60 is an overview of the change in each metal by a change from its 25^{th} to 75^{th} percentile on eGFR values, whilst all other metals are held at a specific percentile. Cadmium is shown to have a significant negative association with eGFR, which appears to be stronger at the lower percentiles (25^{th} and 50^{th}), and when all other metals are held at their respective levels.

To investigate potential nonlinearity of the exposure-response function illustrated in figure 60, I modelled univariate exposure-response functions of each heavy metal on eGFR (figure 61). These models showed interesting results, specifically in the case of Cadmium and Arsenic which appear to have a nonlinear effect. Arsenic however has an inverted U-shaped curve with high confidence intervals at lower concentrations of Cadmium. Barium and Uranium also show slight U-shaped curves; however, these metals likely have a linear relationship as their curves have a similar trajectory with increasing concentrations (although this was not a significant result in these analyses). To ascertain nonlinearity between metals, I modelled the bivariate exposure-response functions of the metal mixtures, and eGFR which is shown in figure 62. The plot shows the exposure-response function for each metal held at its median value (top panel) when its paired metal is held at either the 25^{th} 50^{th} or 75^{th} percentile (right panel). Results suggest interaction between Cadmium and Arsenic in their respective associations with eGFR, as the positive slope for Cadmium gets steeper at lower levels of Arsenic. Due to the large

degree of overlap between confidence intervals however, this is likely not a statistically significant interaction.

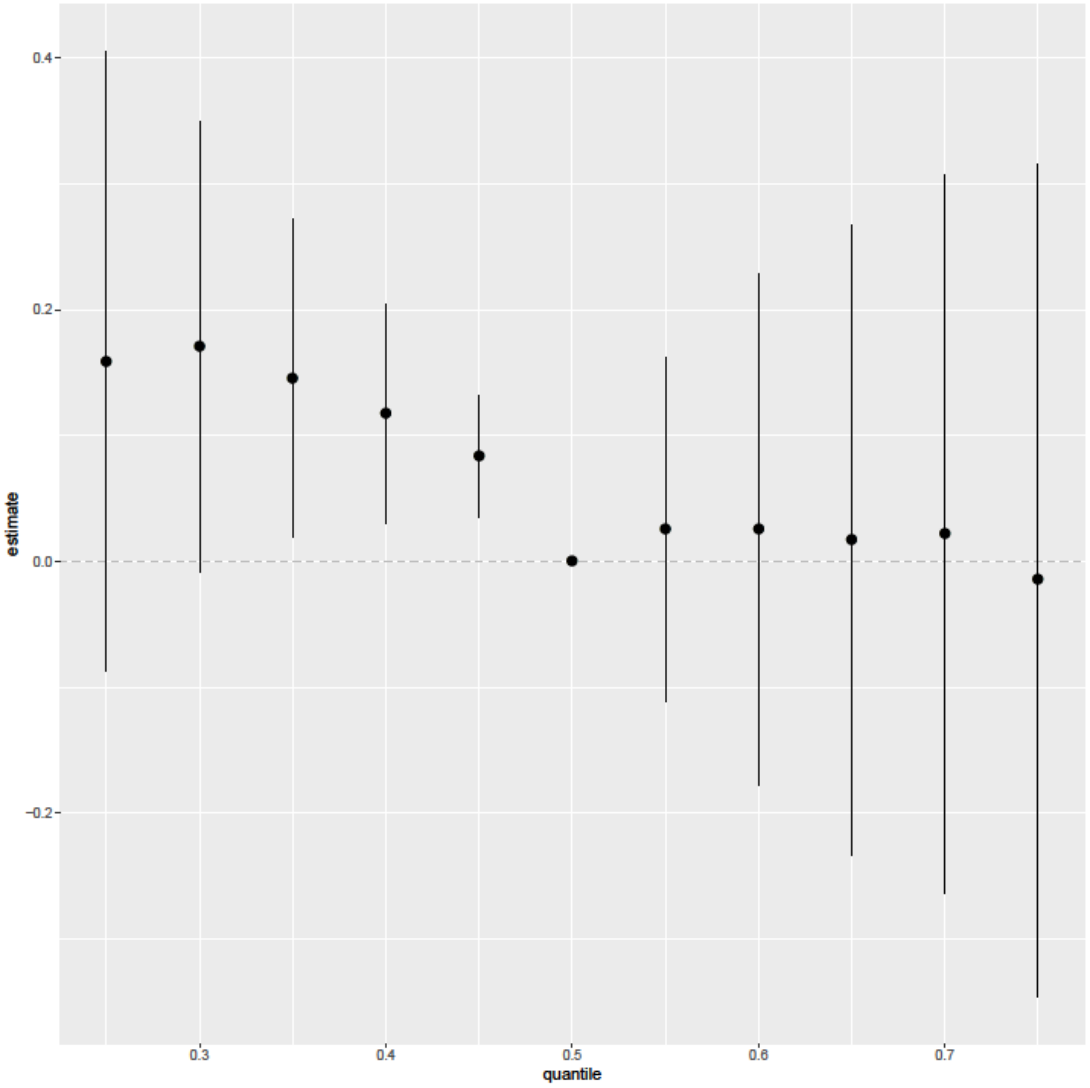


Figure 59. Overall effect of the blood heavy metals mixture on eGFR (y axis) in Bonje when metals are at a specific percentile (shown on the x axis) in comparison to then exposures are all held at the 50th percentile

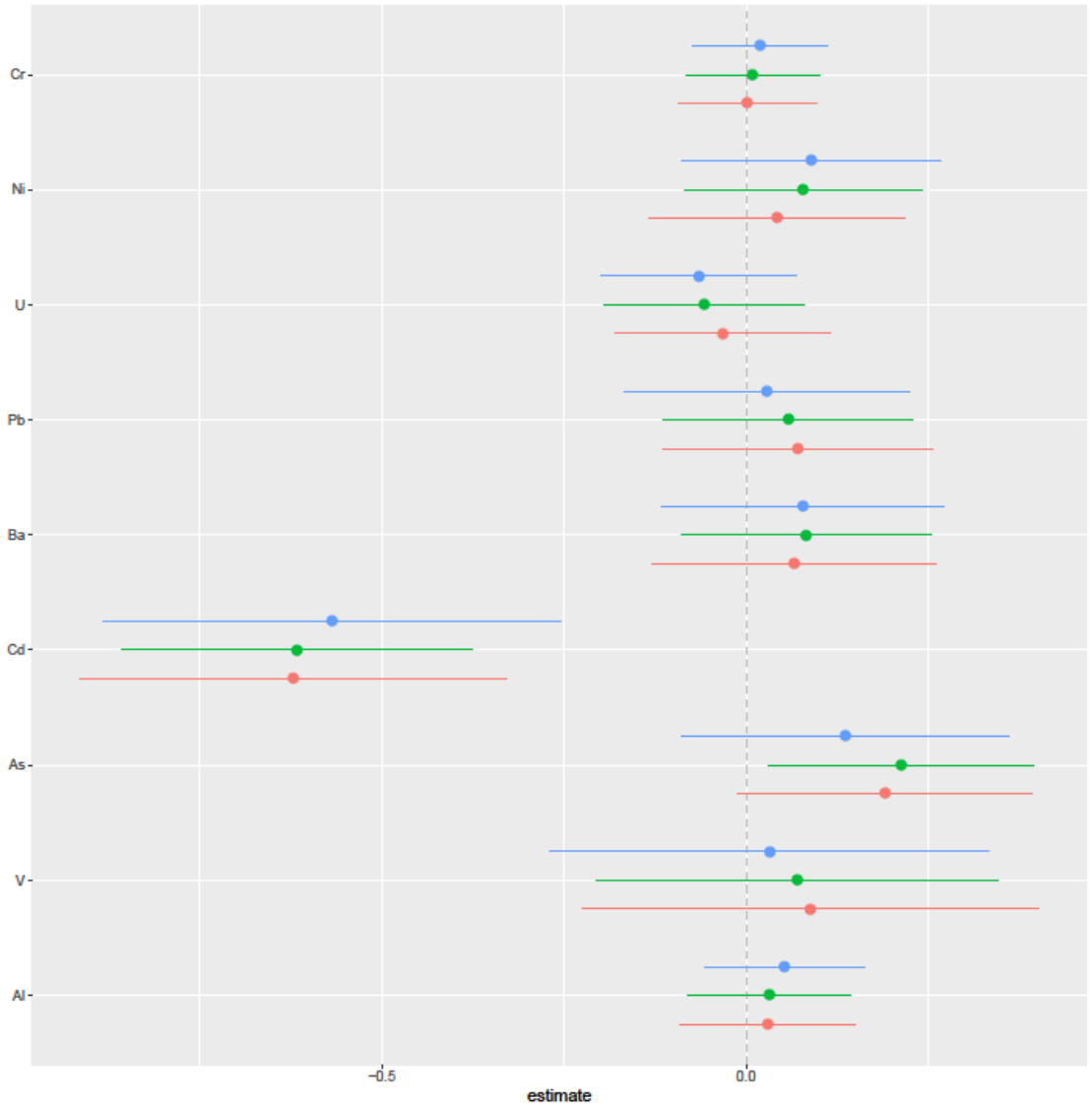


Figure 60. Single metal association on eGFR (x axis) for each metal at its 75th versus 25th percentile, when all other metals are held at either the 25th (red), 50th (green) or 75th (blue) percentile (y axis) in Bonje. Plot shows estimates (dots) and 95% credible intervals (bars), the grey dotted line indicates the null

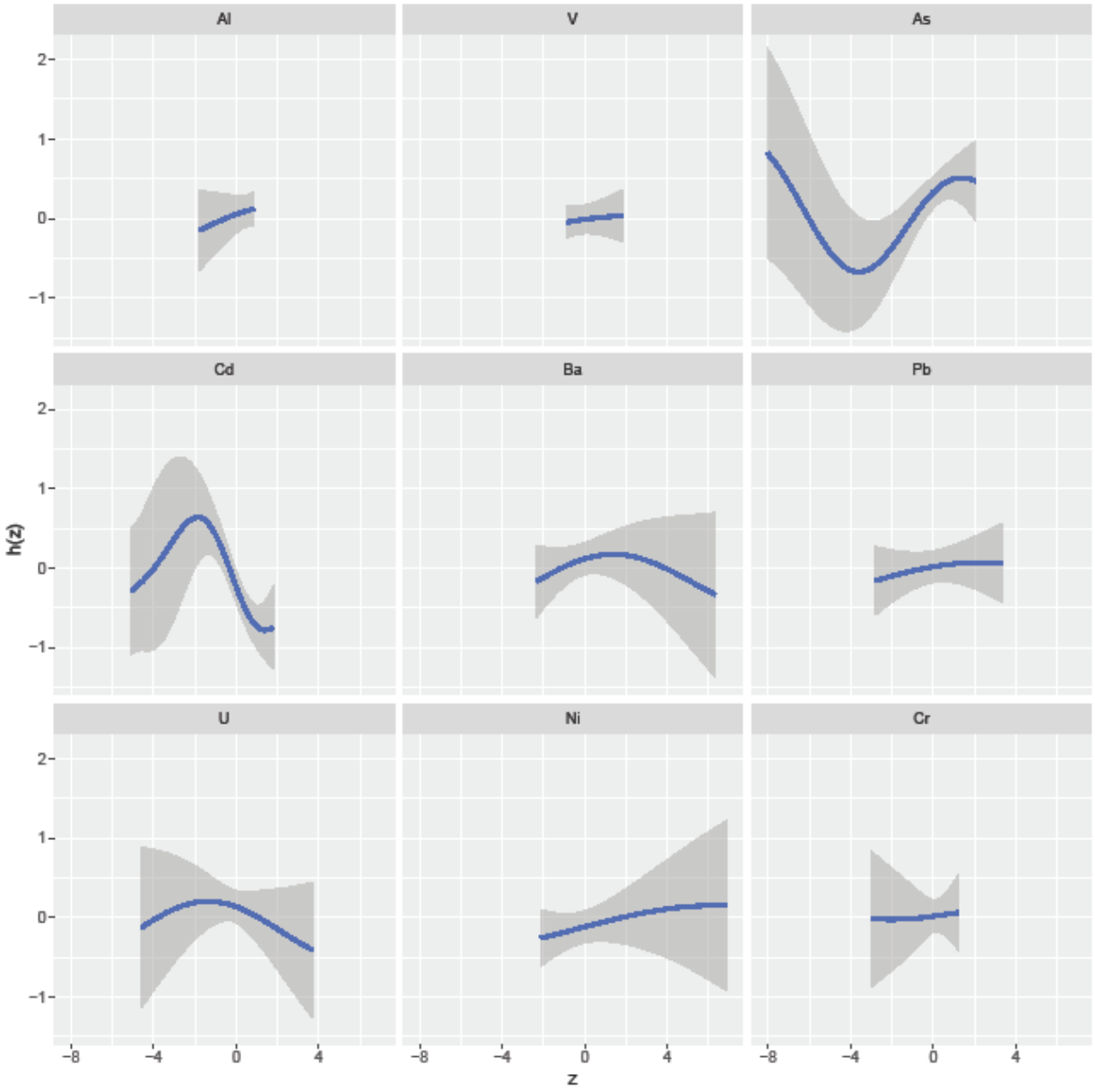


Figure 61. Univariate dose-response functions of heavy metals (x axis) eGFR (y axis) in Bonje. Plot shows 95% confidence intervals for each metal, with all other metals fixed at the 50th percentile

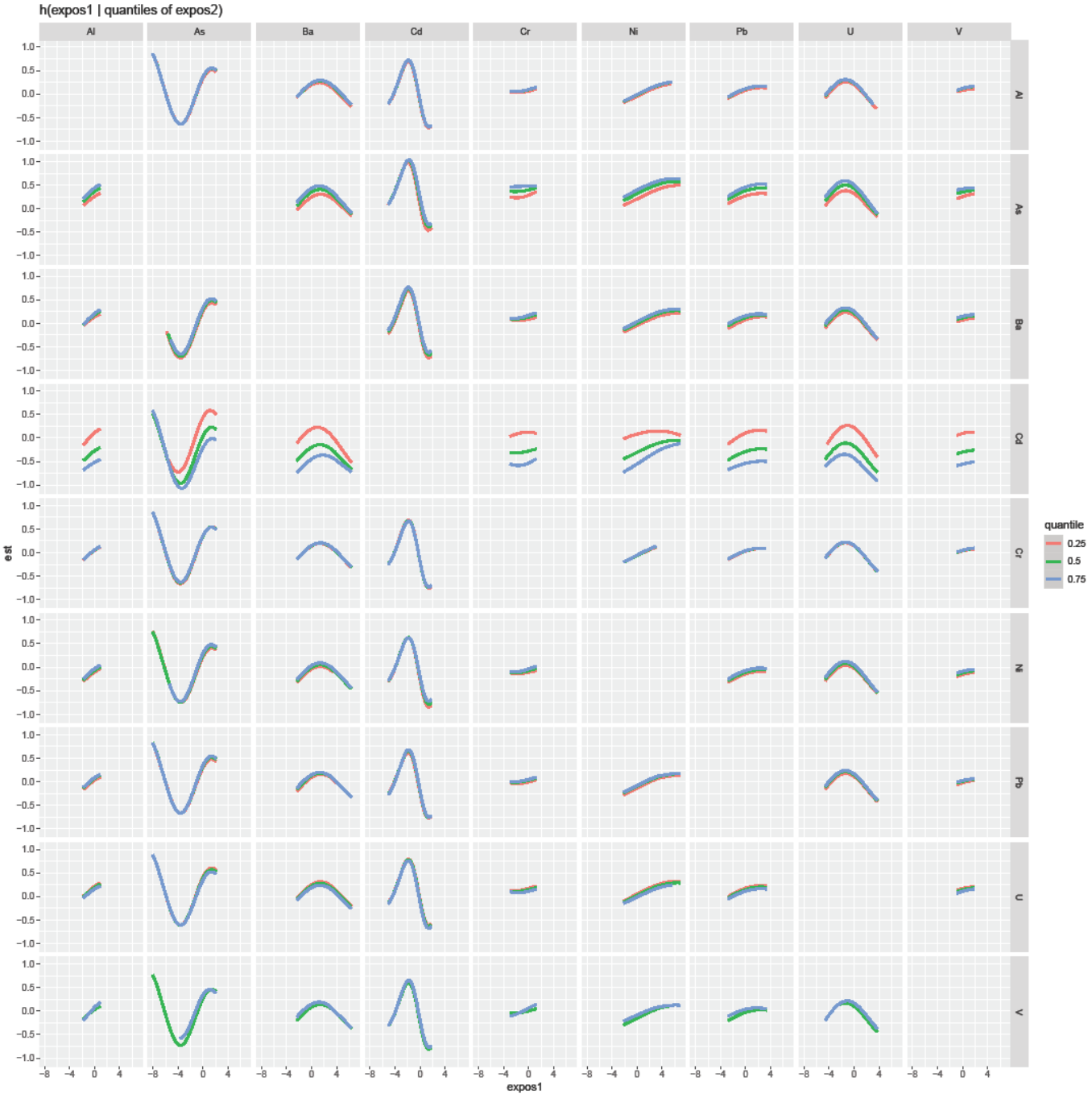


Figure 62. Interactions between metals on eGFR in Bonje. Bivariate exposure-response functions for the metal on the x axis when the metal on the y axis is either at the 25th (red), 50th (green), or 75th (blue) percentile

In Area 25, there was no significant overall effect of the heavy metals' mixture on eGFR (figure 63), however it would appear that at the lower and upper levels of metals (25^{th} and 70^{th}), there is a positive effect on eGFR. The univariate metals overview (Figure 64) shows that at the 25^{th} and 50^{th} percentiles Nickel has a significant negative effect on eGFR, but not at the 75^{th} percentile. There is also a negative effect of Chromium on eGFR at all quantiles, however, this association is not significant. The remainder of the metals in the mixture showed a positive effect on eGFR, however none of these effects were significant. The univariate exposure-response functions of the heavy metals on eGFR (figure 65) show less pronounced results than those seen in Bonje, however, the negative effect of Nickel remains, and is stronger at higher quantiles. There is also a positive effect seen with increasing level of Lead, which could suggest interaction. The bivariate exposure-response functions (figure 66) suggest interaction between Nickel and Lead in their respective associations with eGFR; however, due to the overlap between confidence intervals however, this is likely not a statistically significant interaction.

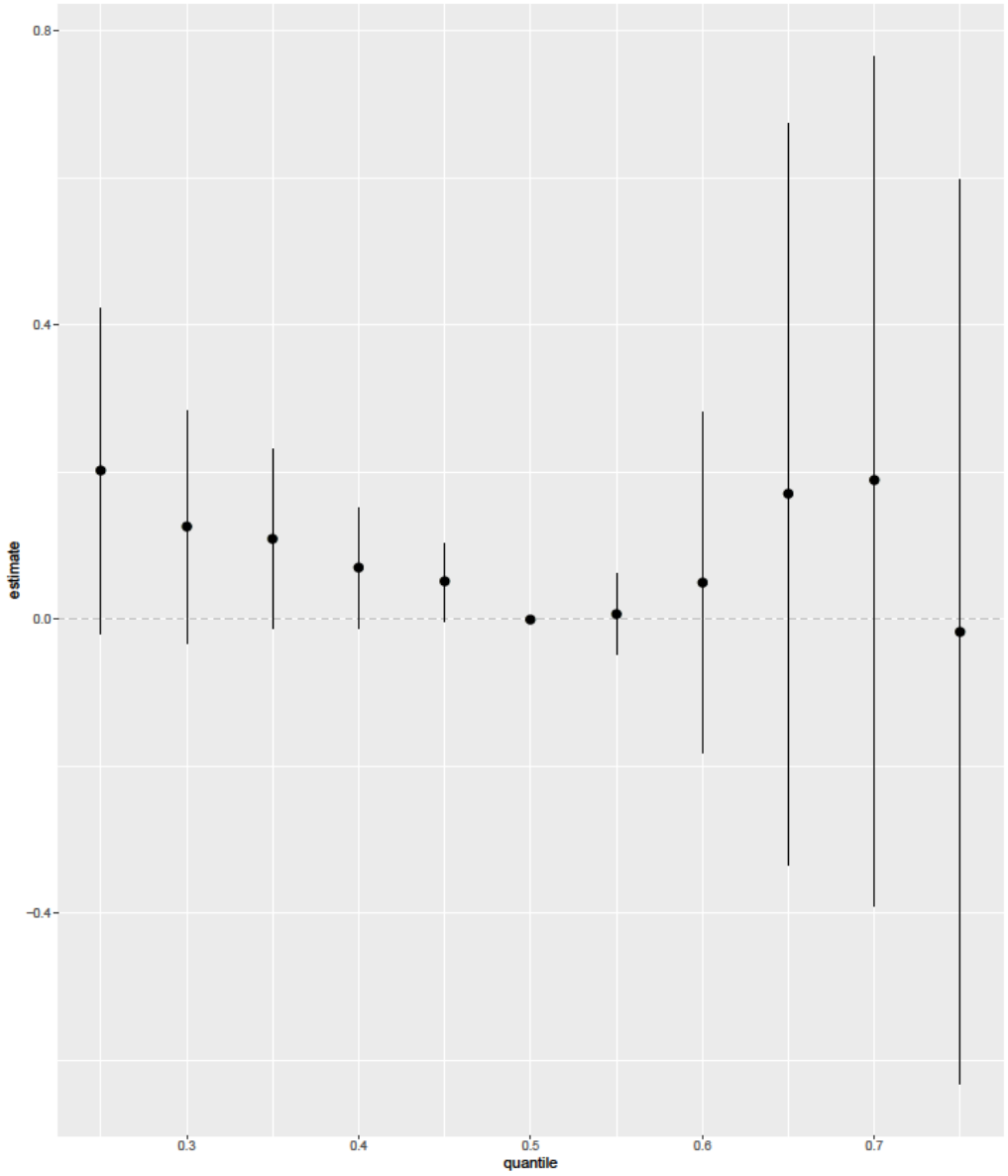


Figure 63. Overall effect of the blood heavy metals mixture on eGFR (y axis) in Area 25 when exposures are at a specific percentile (shown on the x axis) in comparison to then exposures are all held at the 50th percentile

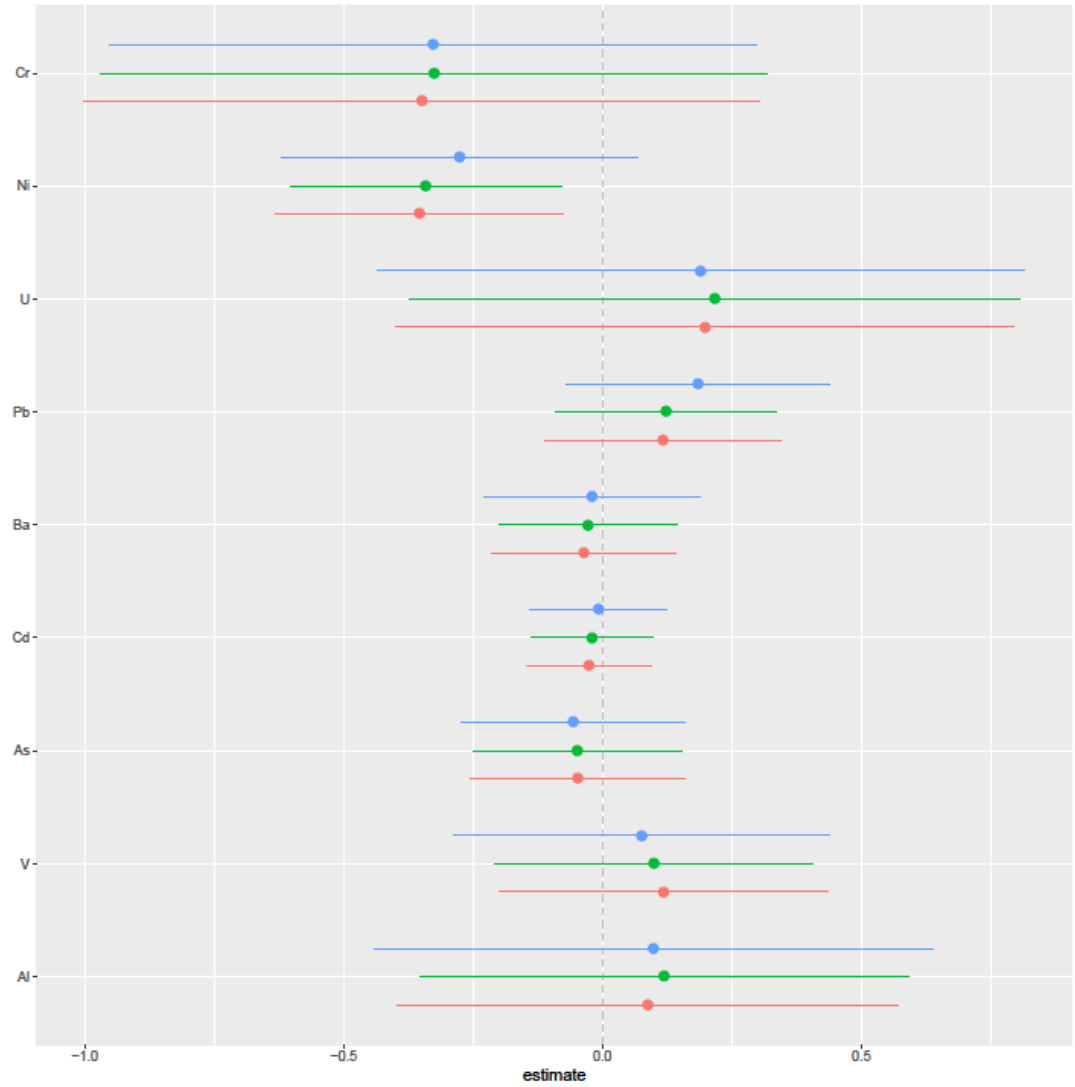


Figure 64. Single metal association on eGFR (x axis) for each metal at its 75th versus 25th percentile, when all other metals are held at either the 25th (red), 50th (green) or 75th (blue) (y axis) in Area 25. Plot shows estimates (dots) and 95% credible intervals (bars), the grey dotted line indicates the null

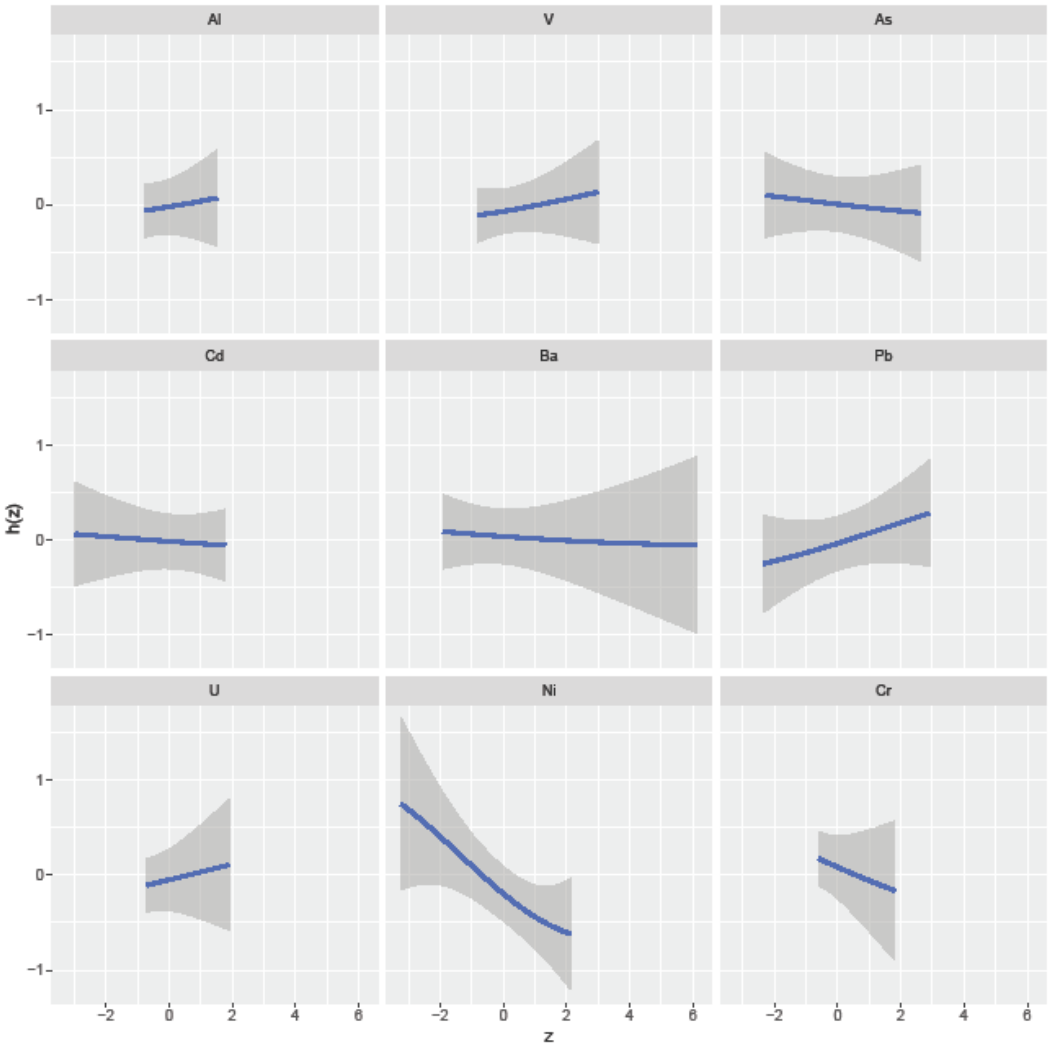


Figure 65. Univariate exposure-response functions of heavy metals (x axis) on eGFR (y axis) in Area 25. Plot shows 95% confidence intervals for each metal, with all other metals fixed at the 50th percentile

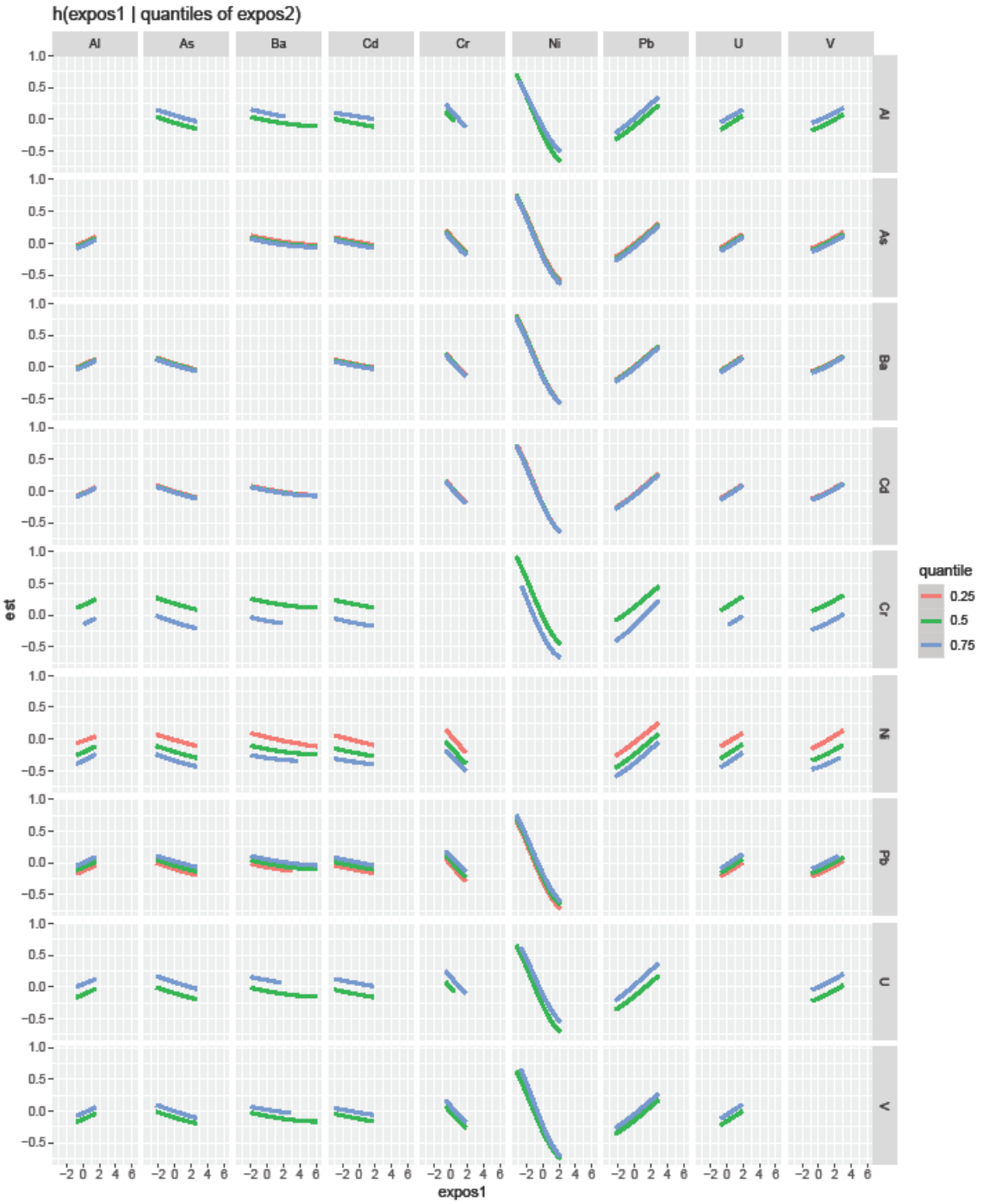


Figure 66. Interactions between metals on eGFR in Area 25. Bivariate exposure-response functions for the metal on the x axis when the metal on the y axis is either at the 25th (red), 50th (green), or 75th (blue) percentile

6.3.6 Water source results

In the following section, I report the water quality of Area 25 and Bonje, characterized by physicochemical properties, and heavy metals concentrations. I then present results from statistical analyses on the distribution of heavy metals per study zone and water source.

The water pH in both Bonje and Area 25 ranged from moderately acidic to alkaline (5.8-7.0; and 6.5-7.7 respectively), and the mean temperature and electrical conductivity were higher in Bonje (28.5± 0.68^{o}C and 726.72± 587.64 S/m respectively) than in Area 25 (24.6±1.76^{o}C and 466.17±385.88 respectively). The mean dissolved oxygen was also higher in Bonje (4.17±2.21 µmol, versus 3.39±1.28 µmol in Area 25), as were turbidity levels (18.35±37.10 NTU, and 2.48±2.23 NTU respectively) (table 50). These observations suggest that the samples in both study zones are under moderately oxidizing conditions. This can cause heavy metals to become mobilized in solution, and have a higher release rate in the water, having a negative impact on the overall water quality[322-324].

Table 50. Descriptive statistics of physicochemical water parameters by water source in Bonje and Area 25

Group	Descriptive statistics	Temperature (°C)	pH	Turbidity NTU	TDS (mg/L)	EC [S/m]	DO (µmol)
				Bonje			
Well							
	Mean	29.1	6.3	9.5	268.5	536.9	3.8
	Median	29.1	6.3	6.8	152.2	304.4	3.2
	Max	29.6	7.0	24.2	902.5	1805.0	7.7
	Min	28.5	5.8	1.3	59.5	119.0	1.2
Borehole							
	Mean	28.8	6.6	27.2	458.3	916.5	4.3
	Median	29.0	6.7	3.1	357.0	714.0	3.5
	Max	30.2	6.9	153.0	875.7	1751.3	2.0
	Min	27.3	6.2	0.2	6.8	13.7	7.9
				Area 25			
Well							
	Mean	24.8	6.8	3.7	375.7	740.0	2.2
	Median	24.6	6.8	3.1	351.1	691.3	2.1
	Max	28.7	7.2	8.9	760.2	1498.7	3.0
	Min	20.3	6.5	0.8	72.7	142.7	1.6
Kiosk							
	Mean	24.6	7.3	1.3	97.7	192.3	4.6
	Median	24.7	7.4	1.4	96.5	190.1	4.5
	Max	26.6	7.7	2.1	107	211.3	5.1
	Min	22.8	7.0	0.5	93.3	183.0	4.1

NTU = Nephelometric Turbidity Units; TDS = total dissolved solids; EC = electrical conductivity; DO = dissolved oxygen

6.3.7 Distribution of heavy metals and trace metals

Overall, the mean heavy metals concentrations did not exceed WHO or MBS permissible limits. In Bonje, the mean Arsenic concentration was 0.11µg/L (range 0.004-0.30µg/L) which was well below the permissible limit. Mean Uranium and Lead concentrations were also below permissible limits at 0.28 µg/L (range 0.003-2.20 µg/L) and 0.49 µg/L (range 0.005-2.30 µg/L) respectively (Table 51). Similarly mean Cadmium concentrations were below permissible limits (0.18 µg/L, range 0.002-2.66 µg/L), however, there was a point source (site 7, covered well, refer to figure 52) which had a Cadmium concentration of 2.66 µg/L, which nears the limit of 3 µg/L[(127)].

Table 51. Summary of heavy metals concentrations across all water sources in Bonje

Metal	n	Mean	Min	Max	Median	SD	WHO limit (µg/l)	MBS limit (µg/l)	Exceedance of WHO, MBS (% of samples)
As	20	0.11	0.00	0.30	0.08	0.08	10	10	0
Cd	20	0.17	0.00	2.66	0.00	0.60	3	3	0
Pb	20	0.49	0.01	2.3	0.25	0.56	10	10	0
Ba	20	3.12	1.12	5.72	3.35	1.20	2400	-	0
U	20	0.28	0.00	2.20	0.04	0.50	30	30	0
V	20	1.53	0.03	6.10	0.79	1.92	-	100	0
Al	20	12.78	1.43	37.70	9.63	11.18	-	200	0
Cr	20	0.30	0.07	0.87	0.30	0.18	50	50	0
Ni	20	0.20	0.05	0.59	0.20	0.15	70	50	0
							(Mg/l)		
$CaCO_3$	20	2.74	0.36	9.51	1.91	2.36	200	-	0
F^-	20	BDL	BDL	BDL	BDL	0	1.5	1.00	0

BDL = below detectable limits

In Area 25 the mean Arsenic, Cadmium, Uranium and Lead levels were below permissible limits (0.09 µg/L, 0.02 µg/L, 1.10 µg/L and 0.27 µg/L respectively (Table 52). For hardness however, approximately 20% of the samples were classified as being “extremely hard”. These high concentrations were seen in three wells (sites 1,3,4) and one kiosk (site 2) (refer to figure 52). There was also one water kiosk (site 20) which had an Aluminium concentration above the permissible limit of 200µg/L (237.95 µg/L).

Table 52. Summary of heavy metals concentrations across all water sources in Area 25

Metal	n	Mean	Min	Max	Median	SD	WHO limit (µg/l)	MBS limit (µg/l)	Exceedance of WHO, MBS (% of samples)
As	20	0.09	0.01	0.19	0.07	0.10	10	10	0
Cd	20	0.02	0.00	0.06	0.00	0.20	3	3	0
Pb	20	0.27	0.09	0.77	0.20	0.15	10	10	0
Ba	20	4.50	2.96	6.07	4.40	1.05	2400	-	0
U	20	1.10	0.01	15.10	0.00	3.36	30	30	0
V	20	1.97	0.42	8.23	1.20	1.97	-	100	0
Al	20	28.73	4.14	237.95	15.70	51.20	-	200	5
Cr	20	0.30	0.05	1.92	0.10	0.47	50	50	0
Ni	20	0.96	0.10	5.06	0.50	1.13	70	50	0
							(Mg/l)		
$CaCO_3$	20	128.85	12.22	541.26	64.90	128.59	200	-	20
F^-	20	BDL	BDL	BDL	BDL	-	1.5	1	0

BDL – below detectable limits

6.3.8 Statistical significance tests for different sampling groups

Table 53 shows the outcome of the Kruskal-Wallis test group significance of the median heavy metals' concentrations between Bonje and Area 25, boreholes and wells, and water kiosks and well. Alongside the Kruskal Wallis results are the post hoc Bonferroni corrections which showed concentrations of Barium, Nickel and hardness were higher in Area 25 than in Bonje. For the Bonje Borehole-Well pair, Nickel was significantly higher in wells; and for Area 25 Arsenic, Barium, Chromium, Nickel, and hardness concentrations were higher in wells.

Table 53. Kruskal Wallis test for heavy metals and pairwise associations of heavy metals concentrations between different water sources (µg/L) within and across Bonje and Area 25

Heavy metal (µg/L)	K-W group significance (Bonferroni correction) *p=<0.05*		
	Bonje-Area 25	Bonje: Borehole-Well	Area 25: Kiosk-Well
As	0.77 (0.31)	0.28 (0.29)	0.01 (0.05)*
Cd	0.01 (0.40)	0.04 (1.00)	0.79 (1.00)
Pb	0.53 (0.10)	0.50 (0.71)	0.19 (1.00)
Ba	0.01 (0.01)***	1.00 (1.00)	0.01 (0.01)***
U	0.01 (0.29)	0.93 (1.00)	0.11 (0.27)
Cr	0.05 (0.09)	0.65 (0.50)	0.05 (0.05)*
V	0.08 (0.47)	0.41 (1.00)	0.01 (0.15)
Al	0.09 (0.19)	0.09 (1.00)	0.09 (1.00)
Ni	0.001 (0.05)*	0.01 (0.05)*	0.05 (0.05)*
CaCO3 (Mg/L)	0.01 (0.01)***	0.49 (1.00)	0.01 (0.01)***

6.3.9 Multidimensional measures of heavy metals

PCA was performed on the correlation matrix of heavy metal data, to identify a reduced set of factors that could capture the variance of metals in water sources across Area 25 and Bonje. I did not model study zones separately due to the small sample size in each area. The scree plot (figure 67) shows that PCs one to four had eigenvalues greater than 1, however the Kaiser criterion[325] was not used in this case, as PCs three and four were increasingly difficult to interpret and did not add value to the results. PCs one and two were selected using the optimal coordinates from the scree plot.

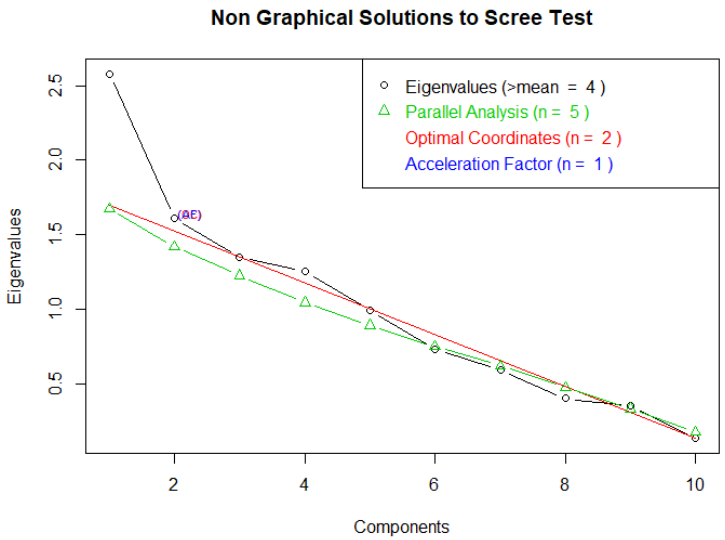


Figure 67. Scree plot showing the cumulative variance explained (y axis) by each principal component (x axis).

Figure 68 is a heat map showing the correlation of heavy metals with each PC which exceeded permissible limits and/or were of a priori interest. PCs one and two accounted for 42% of the cumulative proportion of variance. The first PC (now referred to as RC due to varimax rotation) shows a strong correlation between hardness and Uranium, with a slightly weaker covariance between Vanadium and Nickel in the absence of/or very low concentrations of Arsenic. The second RC shows a strong covariance of Aluminium and Barium in the absence and/or very low concentrations of Nickel. The eigenvalues and cumulative variance for RCs 1-2 are shown in table 54.

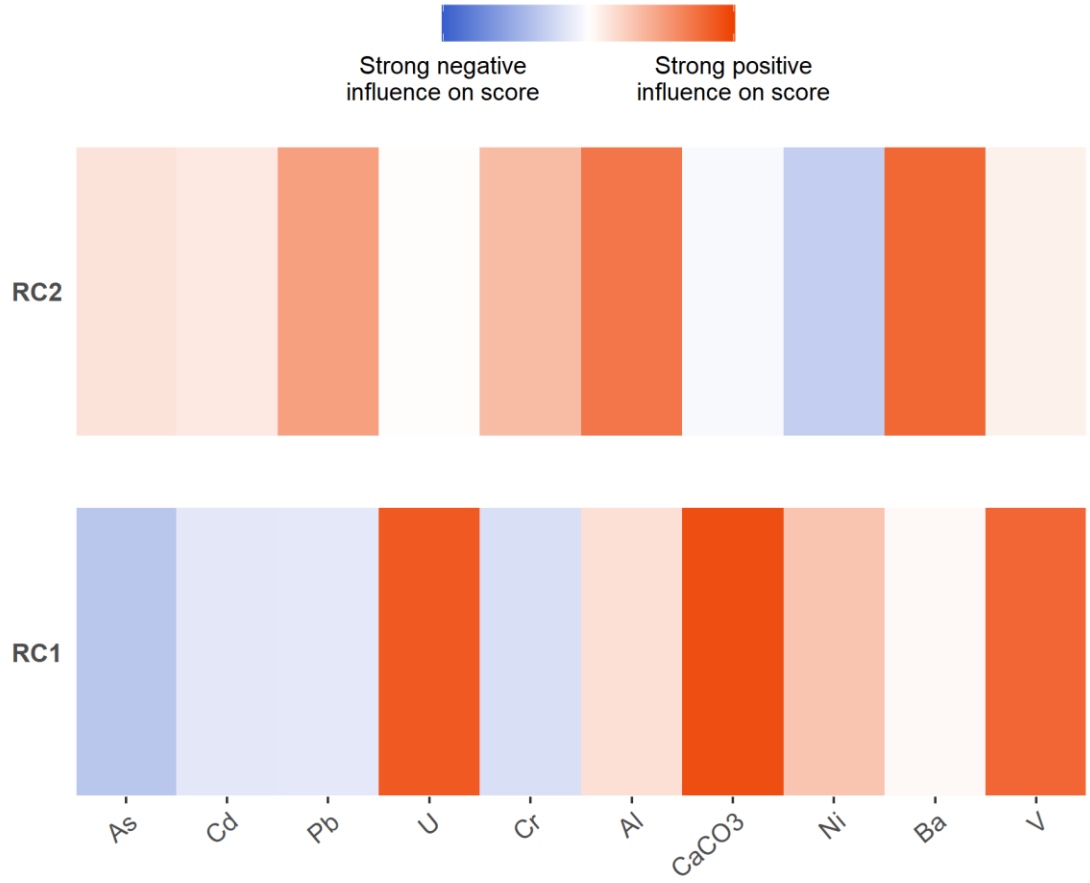


Figure 68. Heat map showing the correlation of heavy metals in water sources across Bonje and Area 25

Table 54. Eigenvalues and associated indicators for PCA analysis on heavy metals

RC	Eigenvalue	Cumulative proportion of variance
RC1	2.58	0.26
RC2	1.61	0.42

To further investigate these PCA results and try to determine which metals could be originating from Bonje versus Area 25 (as the sample size was too small to model the metals by study zone separately), I conducted a t-test on the RC scores from the above PCA analysis to assess whether the metals with highest and lowest influence on the covariance were higher in Bonje or Area 25. The heavy metals scores for RC 1 was statistically significant, suggesting that Area 25 could generally have higher concentrations of the metals which were positively loaded on RCs 1, whereas Bonje may have a higher concentration of those metals which are negatively loaded on the RC1 scores (table 55).

Table 55. T test of varimax-rotated principal component (RC) scores to assess the mean difference in heavy metals concentrations between Bonje and Area 25 across all water sources

RC	Bonje	Area 25	T score	*p* value
	Mean heavy metal concentration			
RC 1	-0.40	0.40	t= -2.81	0.01*
RC 2	-0.04	0.04	t=-0.26	0.78

6.3.10 Blood and water metals combined

To measure the strength of association between metals in blood and water, I modelled a Pearson Correlation matrix. I then conducted linear regression to measure associations between metals in blood, water and eGFR. Hardness was not included in these analyses, as it is not possible to detect this compound in bio samples.

Figure 69 illustrates the water sources participants were assigned to in both study zones using the buffer and spatial join tool. Household locations could not be mapped for data protection purposes.

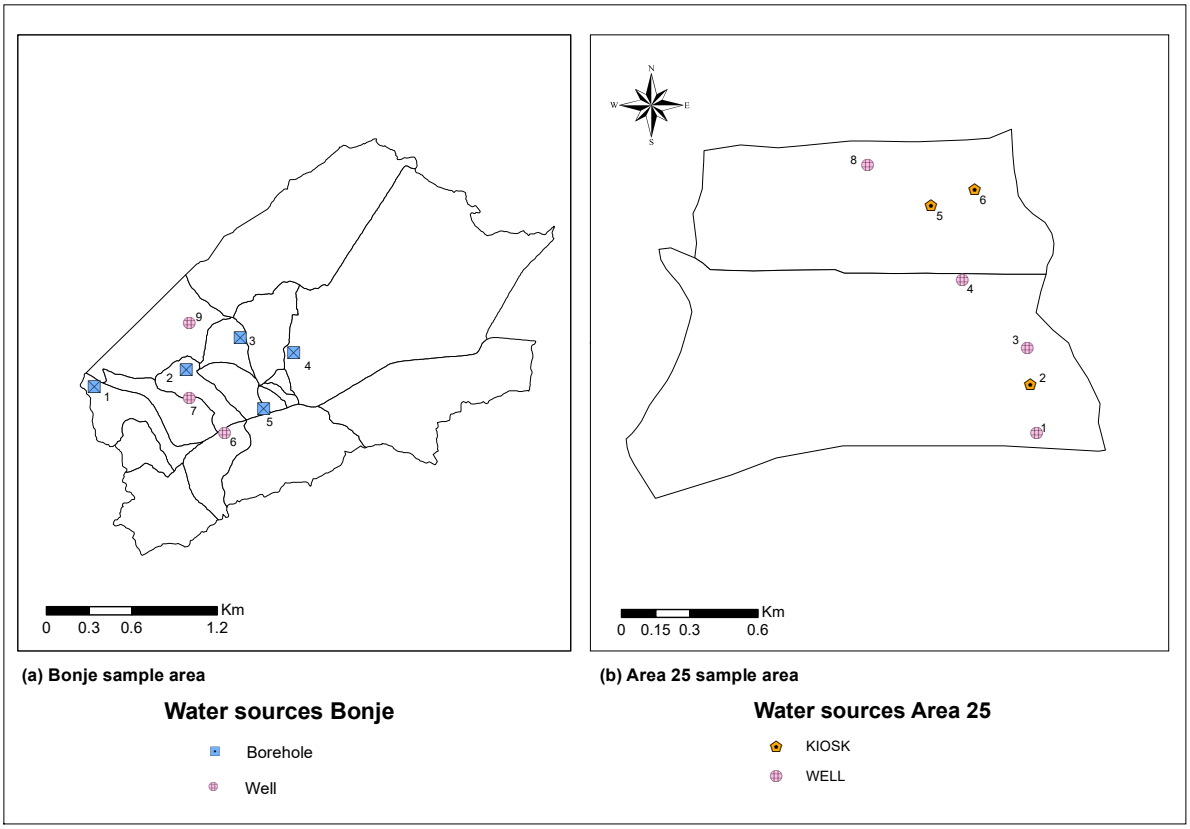


Figure 69. Water sources assigned to participants using (a) 500m buffers across Bonje and (b) 200m buffers across Area 25

The correlation matrix for Bonje (figure 70), indicate a significant negative correlation between blood Chromium and water Vanadium (-0.46, p=0.05), Barium (-0.43, p=0.05) and Nickel (-0.5 p=0.01). Similar patterns were seen with blood Uranium concentrations, but these associations were weaker. In Area 25 negative correlations were observed between Arsenic in blood and water (-0.34, p=0.05) (figure 71).

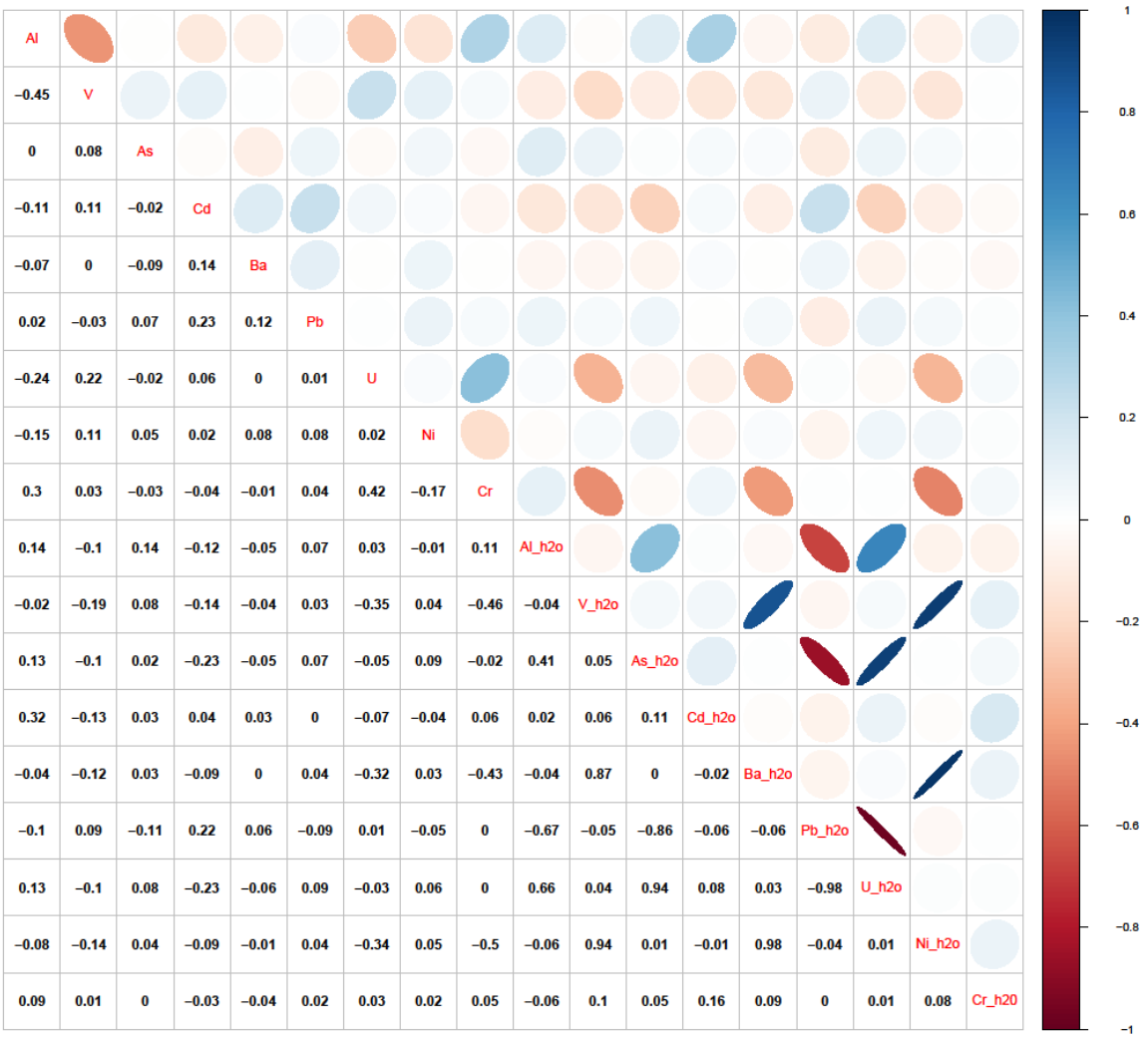


Figure 70. Correlation matrix of blood and water heavy metals in Bonje. Metals labelled "_h2o" denote water metals and those without denote blood metals. Narrower ellipses indicate stronger correlations, and the ellipses' orientation indicates the correlation direction. Cells display significant correlations ($p<0.05$)

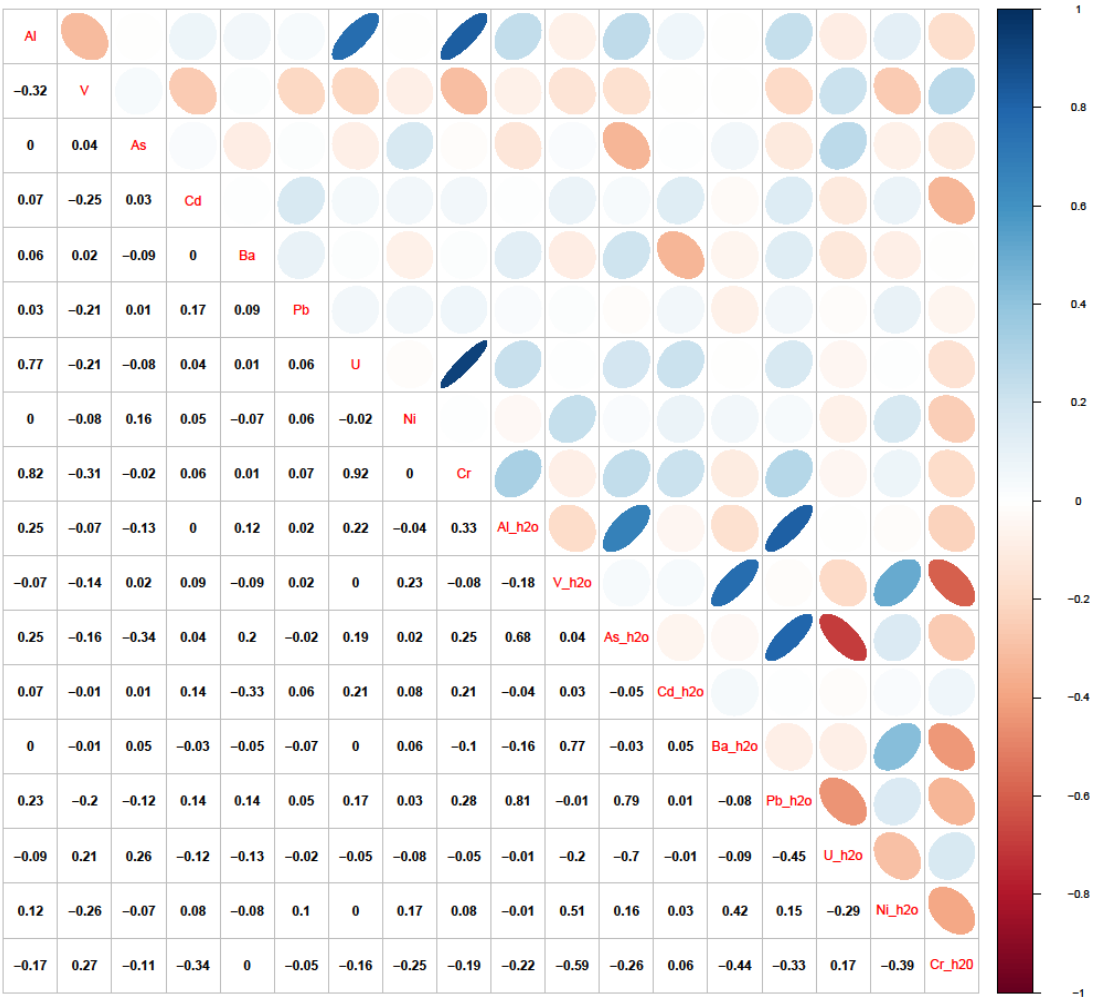


Figure 71. Correlation matrix of blood and water heavy metals in Area 25. Metals labelled "_h2o" denote water metals and those without denote blood metals. Narrower ellipses indicate stronger correlations, and the ellipses' orientation indicates the correlation direction. Cells display significant correlations (p<0.05)

In order to assess whether blood and water metals were associated with eGFR, I conducted linear regression for Bonje and Area 25, adjusting for age, sex, and smoking (table 56). In Bonje a weak negative association was observed with blood Uranium concentrations and eGFR (-0.99 95% CI= -1.95, -0.02). In Area 25 there was a significant negative association between blood and water Chromium concentrations and eGFR (-7.63 95%CI=-14.42, -0.84, and -3.02 95%CI=-5.87, -0.16 respectively). Well water had a significant negative association with eGFR in Area 25 (-7.82, 95%CI=-15.59, -0.05).

Table 56. Linear regression model of metals in blood and water and water source in association with eGFR

Variable	Bonje		Area 25	
	eGFR coefficient	95% CI	eGFR coefficient	95% CI
Blood metals (Log_e)*				
Cd	-0.46	-1.34, 0.41	-2.06	-3.81, -0.31
V	0.22	-0.68, 1.14	1.44	-0.71, 3.60
As	0.64	-0.18, 1.47	1.14	-0.97, 3.26
Ni	-0.14	-0.96, 0.66	-1.80	-1.84, 7.47
Ba	0.20	-0.59, 1.01	0.02	-5.79, 1.97
Pb	-0.70	-1.54, 0.12	0.46	-1.56, 2.48
Cr	-0.48	-0.50, 1.11	-7.63*	-14.42, -0.84
Al	0.71	-0.38, 1.80	1.07	-2.46, 4.61
U	-0.99*	-1.95, -0.02	7.06	0.58, 13.54
Water metals (Log_e)*				
Cd	-0.40	-1.31, 0.49	0.05	-1.75, 1.76
V	0.42	-3.78, 4.63	0.42	-10.07, 10.93
As	16.41	14.45, 18.36	13.14	2.36, 23.92
Ni	-7.11	-17.46, 3.23	-1.80	-1.84, 7.47
Ba	6.13	-0.62, 12.89	-1.90	-5.79, 1.97
Pb	15.30	13.04, 17.57	16.26	2.55, 29.98
Cr	0.30	-0.50, 1.11	-3.02*	-5.87, -0.16
Al	10.31	8.81, 11.82	-30.01	-52.09, -7.93
U	0.77	-0.89, 1.24	15.27	3.60, 26.93
Well*	16.37	-1.67, 34.42	-7.82*	-15.59, -0.05

*Adjusted for age sex and smoking and water source

6.3.11 Modelling the harmonized dataset

Having characterized the blood and water metals in Bonje and Area 25, and analysed these matrices in relation to eGFR separately, I conducted regression analyses on the entire environmental study dataset. As the blood and water had been analysed separately with eGFR alone, and in the absence of other variables, I included all variables in the initial profiling to investigate whether there were any confounding relationships between blood, water and covariates which were not previously modelled.

For these analyses I did not include the heavy metals principal components due to the potential misclassification of water sources which were assigned to participants (due to the limitations discussed above). I was therefore not confident that the PC's would produce reliable results when modelled in association analyses with eGFR and the other covariates.

Furthermore, due to sparse data, I conducted variable profiling of all covariates in association with eGFR in linear and logistic regression models (tables 57-60) which were minimally adjusted for age and sex, (model 1), incrementally adjusted for environmental variables (models 3-5 respectively) and then mutually adjusted for all covariates (model 6 respectively). As part of the sensitivity analysis, I firstly modelled covariates with heat index derived from EBK interpolation (tables 57 and 58), followed by models which were created using heat index from bilinear interpolation models (tables 59 and 60). The results from these models highlight the issues with sparse data, which can be seen in the small coefficients and wide confidence intervals. For this reason, I report result for the linear regression models only (models 6 tables 57 and 59).

In the minimally adjusted model which uses EBK data (table 57 model 1), altitude was the only satellite derived environmental variable which had a negative association with eGFR [regression coefficient (95%CI) = -5.04 (-6.65, -3.42)], whereas crop cover [regression coefficient (95%CI) = 5.61 (0.79, 10.44)], and HI had a positive association [regression coefficient (95%CI) = 11.95 (8.24, 15.67)]. Amongst the environmental questionnaire variables, the use of wells and kiosks for potable water [regression coefficient (95%CI) = -5.58 (-8.77, -2.38), and -10.50 (-13.19, -7.87) respectively], had the strongest negative association with eGFR.

Furthermore, a negative association was observed with blood aluminium concentration [regression coefficient (95%CI) = -1.13 (--1.03, -0.23)]. Finally, those identifying as small holder farmers had a positive association with eGFR (table 57).

In the fully adjusted model (6), there was multicollinearity between increasing altitude and heat index, therefore these variables were modelled separately. Furthermore, the variance inflation factors between study zone, altitude and heat index values were nearing the threshold (VIF=7 and 8 respectively) therefore I decided to remove study zone from the final models in order to understand the effect of environmental risk factors on eGFR without potential inflation risk from the zone variable.

In this model, age and blood Al concentrations maintained a negative association with eGFR, and small holder farmers maintained a positive association. No associations between environmental variables and eGFR remained.

Table 57. Associations between sociodemographic, anthropometric, and environmental characteristics and eGFR in Malawian participants using heat index data from empirical Bayesian kriging, n=367

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]	Model 7[g]
	Coefficient (95%CI)						
Age (10-year increase)[h]	-8.57***	-8.82***	-8.82***	-8.82***	-8.85***	-8.83***	-8.85***
	(-9.35, -7.79)	(-9.72, -7.93)	(-9.72, -7.92)	(-9.72, -7.92)	(-9.74, -7.95)	(-9.72, -7.93)	(-9.74, -7.95)
Sex [i]							
Male	0.82	0.52	0.53	0.52	0.54	0.45	0.53
	(-1.46, 3.11)	(-2.15, 3.17)	(-2.12, 3.19)	(-2.13, 3.18)	(-2.09, 3.18)	(-2.13, 3.11)	(-2.12, 3.18)
Female	REF	REF	REF	REF	REF	REF	REF
Water source							
Well	-5.58***	-4.08*	-4.08*	-4.08*	-3.59	-4.07*	-3.62
	(-8.77, -2.38)	(-7.92, -0.24)	(-7.93 -0.23)	(-7.93, -0.23)	(-7.48, 0.28)	(-7.91, -0.20)	(-7.53, 0.29)
Kiosk	-10.50**	-5.79	-7.35	-6.06	1.30	2.39	2.29
	(-13.19, -7.81)	(-15.74, 4.16)	(-15.74, 5.50)	(-19.07, 6.93)	(-10.77, 3.77)	(-19.07,7.87)	(-5.13, 19.72)
Borehole	REF	REF	REF	REF	REF	REF	REF
Muscle cramps							
Yes	0.57	-1.51	-1.48	-1.50	-1.63	-1.57	-1.63
	(-1.77, 2.92)	(-4.27, 1.24)	(-4.26, 1.29)	(-4.28, 1.26)	(-4.39, 1.12)	(-4.34, 1.19)	(-4.40, 1.13)
No	REF	REF	REF	REF	REF	REF	REF
Farming category							
Commercial	REF	REF	REF	REF	REF	REF	REF
Small holder	4.97*	4.34*	4.33*	4.34*	4.32*	4.37*	4.32*
	(1.18, 8.76)	(0.39, 8.29)	(0.37, 8.29)	(0.38, 8.30)	(0.38, 8.25)	(0.38, 8.32)	(0.37, 8.27)
Subsistence	-0.03	0.05	0.05	0.04	0.25	0.13	0.25
	(-3.09, 3.02)	(-2.96, 3.06)	(-3.04, 3.06)	(-3.00, 3.09)	(-2.76, 3.27)	(-3.00, 3.16)	(-2.77, 3.28)
Blood							
Al	-1.13	-1.61*	-1.58*	-1.61*	-1.62*	-1.65*	-1.63*
	(-1.03, 1.23)	(-3.04, -0.18)	(-3.02, -0.14)	(-3.04, -0.17)	(-3.04, -0.20)	(-3.08, -0.22)	(-3.06, -0.20)
Water							
Cd	0.37	0.55	0.56	0.56	0.51	0.54	0.52
	(-0.79, 1.53)	(-0.79, 1.81)	(-0.69, 1.83)	(-0.70, 1.82)	(-0.73, 1.77)	(-0.71, 1.81)	(-0.74, 1.78)
Ni	-3.87***	-1.32	-1.31	-1.31	-1.22	-1.33	-1.23
	(-4.93, -2.81)	(-3.83, 1.19)	(-3.83, 1.19)	(-3.84, 1.20)	(-3.73, 1.28)	(-3.85, 1.18)	(-3.75, 1.28)
Study zone							
Bonje	8.93***	-	7.29	-	-	-	-
	(6.59, 11.27)		(-12.79, 9.62)				
Area 25	REF		REF				
Altitude (m)	-5.04***	-	-	0.02	-	-	0.37
	(-6.65, -3.42)			(-0.92, 0.97)			(-0.77, 1.52)

Heat indexEBK (0.5)	11.95** (8.24, 15.67)	-	-	-	1.41 (-0.48, 3.31)	-	1.34 (-0.88, 3.56)
Land cover							
Cropland	5.61* (0.79, 10.44)	-	-	-	-	8.24 (-10.73,9.21)	13.36 (-8.76, 13.48)
Urban	REF					REF	REF

[a] minimal adjustment for age, sex , [b] base variables mutually adjusted; [c] Base variables adjusted for study zone; [d] Base variables adjusted for altitude; [e] Base variables adjusted for heat index, [f] Base variables adjusted for land cover; [g] fully adjusted models; [h] adjusted for age; [i] adjusted for sex; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

Table 58. Associations between sociodemographic, anthropometric, and environmental characteristics and eGFR <90 in Malawian participants using heat index data from empirical Bayesian kriging, n=367

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]	Model 7[g]
			Coefficient (95%CI)				
Age (10-year increase)[h]	5.85***	8.82***	8.71***	8.75***	8.65***	8.72***	8.65***
	(3.89, 14.87)	(3.64, 11.93)	(3.64, 10.92)	(3.65, 12.62)	(4.74, 7.91)	(7.64, 9.24)	(4.95, 10.31)
Sex [i]							
Male	1.82	3.27	3.27	3.27	3.20	3.27	3.15
	(-1.46, 3.11)	(0.58, 4.24)	(0.58, 3.64)	(0.58, 3.62)	(0.54, 2.34)	(0.58, 4.24)	(0.53, 3.32)
Female	REF	REF	REF	REF	REF	REF	
Water source							
Well	2.29**	3.35	3.35	3.30	2.92	3.35	2.91
	(1.4`, 6.74)	(0.23, 6.12)	(0.23 6.12)	(0.23, 6.03)	(0.18, 5.43)	(0.23, 6.12)	(0.18, 5.38)
Kiosk	3.13*	NA	NA	NA	NA	NA	NA
	(3.02, 24.46)						
Borehole	REF	REF	REF	REF	REF	REF	REF
Muscle cramps							
Yes	0.46	0.59	0.59	0.60	0.59	0.59	0.59
	(0.13, 1.58)	(0.08, 4.24)	(0.08, 4.24)	(0.08, 4.33)	(0.07, 4.61)	(0.08, 4.24)	(0.07, 4.75)
No	Ref	REF	REF	REF	REF	REF	REF
Farming category							
Commercial	REF	REF	REF	REF	REF	REF	REF
Small holder	1.78	NA	NA	NA	NA	NA	NA
	(0.17, 33.36)						
Subsistence	2.26	NA	NA	NA	NA	NA	NA
	(0.31, 24.57)						
Blood							
Al	1.33	1.98	1.98	1.96	1.99	1.98	1.98
	(0.73, 2.42)	(0.64, 6.19)	(0.64, 6.15)	(0.64, 6.19)	(0.61, 6.68)	(0.64, 6.15)	(0.61, 6.65)
Water							
Cd	0.57	0.60	0.61	0.60	0.56	0.60	0.56
	(0.24, 1.13)	(0.18, 1.53)	(0.18, 1.53)	(0.18, 1.52)	(0.15, 1.46)	(0.18, 1.53)	(0.16, 1.46)
Ni	2.40	0.87	0.87	0.87	0.63	0.87	0.62
	(1.20, 5.27)	(0.12, 6.19)	(0.12, 6.18)	(0.12, 6.15)	(0.07, 4.79)	(0.12, 6.19)	(0.07, 4.75)
Study zone							
Bonje	0.19*		1.04				
	(0.02, 0.62)		(1.01, 3.62)				
Area 25	REF		REF				
Altitude (m)	1.19*			1.24			NA
	(1.05, 1.40)			(0.01, 2.66)			

Heat index EBK (0.5)	0.63*			0.33	0.32
	(0.44, 0.87)			(0.09, 2.43)	(0.09, 2.39)
Land cover					
Cropland	0.13*			1.04	1.05
	(0.02, 0.75)			(0.08, 7.21)	(0.76, 33.48)
Urban	REF			REF	REF

[a] minimal adjustment for age, sex , [b] base variables mutually adjusted; [c] Base variables adjusted for study zone; [d] Base variables adjusted for altitude; [e] Base variables adjusted for heat index, [f] Base variables adjusted for land cover; [g] fully adjusted models; [h] adjusted for age; [i] adjusted for sex Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

=

6.3.1.2 Modelling data using bilinear interpolation heat index data

I modelled the data using linear regression with the bilinear interpolation heat index data across Malawi.

Overall, results from models which use this heat index data (table 59 models 1-6) were similar to those observed in the previous models shown above. The negative association between age an eGFR remained, however, the association between eGFR and Aluminium in the blood was lost. Furthermore, the protective effect of small-holder farmers was lost.

Table 59. Associations between sociodemographic, anthropometric, and environmental characteristics and eGFR in Malawian participants using heat index data from bilinear interpolation methodology, n=367

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]	Model 7[g]
			Coefficient (95%CI)				
Age (10-year increase)[h]	-8.57*** (-9.35, -7.79)	-8.74*** (-9.99, -5.70)	-8.79*** (-9.94, -4.65)	-8.58*** (-8.94, -4.66)	-8.57*** (-9.94, -6.65)	-8.79*** (-8.94, -4.65)	-8.88*** (-9.02, -4.74)
Sex [i]							
Male	0.82 (-1.46, 3.11)	0.22 (-2.50, 2.96)	0.32 (-2.40, 3.05)	0.30 (-2.42, 3.03)	0.32 (-2.40, 3.05)	0.32 (-2.40, 4.65)	0.70 (-2.02, 3.43)
Female	REF	REF	REF	REF	REF	REF	REF
Water source							
Well	-5.58*** (-8.77, -2.38)	-2.55 (-5.76, 0.65)	-2.51 (-5.71, 0.68)	-2.54 (-5.74, -0.62)	-2.51 (-5.71, 0.68)	-2.51 (-5.71, -0.68)	-1.95 (-3.53, 1.39)
Kiosk	-10.50** (-13.19, -7.81)	-5.17 (-19.36, -1.96)	-6.65 (-20.04, 2.26)	-6.14 (-20.03, 2.15)	-6.15 (-20.04, 2.16)	-6.15 (-20.04,2.26)	1.81 (-2.13, 13.72)
Borehole	REF	REF	REF	REF	REF	REF	REF
Muscle cramps							
Yes	0.57 (-1.77, 2.92)	-1.79 (-4.12, 0.53)	-1.69 (-4.02, 0.63)	-2.54 (-4.03, 0.62)	-1.69 (-4.02, 0.63)	-1.69 (-4.02, 0.63)	-1.86 (-4.17, 0.45)
No	REF	REF	REF	REF	REF	REF	REF
Farming category							
Commercial	REF	REF	REF	REF	REF	REF	REF
Small holder	4.97* (1.18, 8.76)	2.77 (-0.57, 0.65)	2.73 (-0.37, 1.37)	2.73 (-0.61, 6.07)	2.73 (-0.61, 6.07)	2.73 (-0.61, 6.07)	2.81 (-0.54, 6.12)
Subsistence	-0.03 (-3.09, 3.02)	0.97 (-3.54, 1.59)	0.05 (-3.78, 1.37)	0.04 (-3.79, 1.39)	0.25 (-3.78, 1.37)	0.20 (-3.79, 1.37)	0.25 (-3.49, 1.66)
Blood							
Al	0.10 (-1.03, 1.23)	-0.84 (-2.10, 0.41)	-0.72 (-1.99, 0.53)	-0.77 (-2.03, 0.48)	-0.72 (-1.98, 0.54)	-0.72 (-1.99, 0.53)	-0.02 (-1.41, 1.47)
Water							
Cd	0.37 (-0.79, 1.53)	0.61 (-0.49, 1.74)	0.62 (-0.69, 1.83)	0.61 (-0.48, 1.75)	0.61 (-0.48, 1.75)	0.61 (-0.48, 1.73)	0.59 (-4.19, 1.73)
Ni	-3.87*** (-4.93, -2.81)	-0.48 (-2.58, 1.62)	-0.43 (-3.83, 1.19)	-0.44 (-2.51, 1.67)	-0.43 (-2.53, 1.66)	-0.43 (-2.53, 1.65)	-1.23 (-1.90, 2.47)

Study zone							
Bonje	8.93***	-	6.62	-	-	-	-
	(6.59, 11.27)		(-12.79, 9.62)				
Area 25	REF		REF				
Altitude (m)	-5.04***	-	-	0.50	-	-	-1.21
	(-6.65, -3.42)			(-0.31, 1.31)			(-2.32, -0.99)
Heat index bilinear interpolation (0.5 ºC)	2.16**	-	-	-	2.47	-	-1.94
	(1.60, 2.73)				(-0.48, 3.31)		(-3.77, 1.08)
Land cover							
Cropland	5.61*	-	-	-	-	6.62	15.83
	(0.79, 10.44)					(-16.26, 12.93)	(-18.76, 13.94)
Urban	REF					REF	REF

[a] minimal adjustment for age, sex, [b] base variables mutually adjusted; [c] Base variables adjusted for study zone; [d] Base variables adjusted for altitude; [e] Base variables adjusted for heat index, [f] Base variables adjusted for land cover; [g] fully adjusted models; [h] adjusted for age; [i] adjusted for sex; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

Table 60. Associations between sociodemographic, anthropometric, and environmental characteristics and eGFR <90 in Malawian participants using heat index data from bilinear interpolation methodology, n=367

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]	Model 7[g]
				eGFR <60 OR (95%CI)			
Age (10-year increase)[h]	5.85*** (3.89, 14.87)	7.43*** (1.26, 7.89)	7.43*** (1.29, 7.89)	8.28*** (1.61, 8.29)	7.43*** (1.27, 7.93)	9.47*** (1.37, 10.15)	9.58*** (1.67, 10.11)
Sex [i]							
Male	1.82 (-1.46, 3.11)	2.81 (0.01, 5.12)	2.80 (0.01, 5.12)	2.80 (0.02, 3.65)	2.73 (0.01, 5.12)	2.80 (0.50, 2.93)	2.73 (0.47, 4.11)
Female	REF	REF	REF	REF	REF	REF	
Water source							
Well	2.29** (1.4`, 6.74)	2.89 (0.76, 6.79)	2.89 (0.01, 1.12)	1.14 (0.02, 2.98)	2.89 (0.65, 6.83)	2.89 (0.35, 8.37)	3.19 (0.27, 8.35)
Kiosk	3.13* (3.02, 24.46)	NA	NA	NA	NA	NA	NA
Borehole	REF	REF	REF	REF	REF	REF	REF
Muscle cramps							
Yes	0.46 (0.13, 1.58)	0.21 (0.01, 2.19)	0.18 (0.01, 2.19)	0.12 (0.01, 1.70)	0.21 (0.01, 2.21)	0.28 (0.06, 3.14)	0.45 (0.06, 3.25)
No	Ref	REF	REF	REF	REF	REF	REF
Farming category							
Commercial	REF	REF	REF	REF	REF	REF	REF
Small holder	1.78 (0.17, 33.36)	NA	NA	NA	NA	NA	NA
Subsistence	2.26 (0.31, 24.57)	NA	NA	NA	NA	NA	NA
Blood							
Al	1.33 (0.73, 2.42)	1.02 (0.20, 4.57)	1.02 (0.20, 4.57)	1.58 (0.25, 9.22)	1.02 (0.20, 4.57)	1.02 (0.61, 5.61)	1.76 (0.54, 5.50)
Water							
Cd	0.57 (0.24, 1.13)	0.57 (0.01, 3.89)	0.58 (0.01, 3.89)	0.53 (0.01, 4.46)	0.58 (0.01, 3.96)	0.57 (0.01, 4.06)	0.40 (0.01, 6.32)
Ni	2.40 (1.20, 5.27)	0.64 (0.05, 5.50)	0.64 (0.05, 5.55)	0.68 (0.06, 3.01)	0.64 (0.04, 5.56)	0.64 (0.20, 7.18)	0.62 (0.12, 7.76)
Study zone							
Bonje	0.19*		NA				

	(0.02, 0.62)					
Area 25	REF					
Altitude (m)	1.19*		0.12			NA
	(1.05, 1.40)		(0.01, 3.01)			
Heat index bilinear	0.77*			0.84		0.90
interpolation (0.5 ºC)	(0.62, 0.94)			(0.48, 1.01)		(0.01, 1.11)
Land cover						
Cropland	0.13*				0.75	0.53
	(0.02, 0.75)				(0.02, 1.98)	(0.01, 1.40)
Urban	REF				REF	REF

[a] minimal adjustment for age, sex , [b] base variables mutually adjusted; [c] Base variables adjusted for study zone; [d] Base variables adjusted for altitude; [e] Base variables adjusted for heat index, [f] Base variables adjusted for land cover; [g] fully adjusted models; [h] adjusted for age; [i] adjusted for sex; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

6.4 Discussion

In this study, environmental and lifestyle variable, and biomarkers for heavy metals were analysed with eGFR as a continuous variable. Overall, very few variables were associated with low eGFR in this sub sample, and heavy metals in both blood and water did not exceed permissible limits. Through use of exploratory methods such as BKMR and PCA some interesting observations were made, primarily regarding mixtures of heavy metals in blood. General findings will be discussed in detail below.

Bonje is situated at a relatively low altitude, with participant residing at altitudes ranging between 500-531 metres. This population is entirely rural, with all participants living within 20m of "cropland", with a heat index range 21.7-22.2^{o}C. Comparatively, Area 25 is situated at a higher altitude, with participants residing between 1099-1125m, and has a lower heat index range of 19.3-19.4^{o}C. The land cover categories are a mix of "urban" and "cropland". In general, this Malawian population reside at higher altitudes, and cooler temperatures than those reported in CKDu-endemic regions[16, 31, 289] which could contribute towards the generally higher mean eGFR value seen in this population (117.0±16.5 mL/min/1.73m^2). This factor, however, could also be attributed to the young average age of the Malawian sample, both overall (33.5±12.7 years) and the nested sample (36.0±4.4 years).

Like most of the regions in which CKDu has been reported [164, 326, 327], a majority of the population was involved in agricultural work (66%, n=245), with over half of these workers situated in Bonje (58%, n=213). Furthermore, 80% were employed on an annual basis, and reported taking regular rest breaks at work or during their daily activities. The most often reported work-related adverse symptoms were thirst (n=267) and sweat (n=247), however raised heart rate was the least reported symptom suggesting that this population might not be conducting such intensive labour as reported by CKDu endemic communities[11, 13, 181, 328].

Approximately 58% (n=214) of participants answered the agricultural questionnaire section; 98% (n=210) of which were Bonje residents. Half of the farming practices reported were for subsistence purposes (49.5%), followed by commercial (29.9%). In Area 25 the most cultivated crop was maize; whilst in Bonje rice, maize, wheat, and cassava were the most cultivated, largely on a subsistence basis. Approximately 10% (n=22) of the agricultural questionnaire

respondents reported self-application of pesticides, of which very few reported use of PPE when applying pesticides, which included facemasks (n=3), overalls (n=4) and footwear (n=3). Overall, across each study zone the mean heavy metal blood concentrations did not exceed permissible limits; however, concentrations were marginally higher in Bonje. In both zones there were some participants with blood concentrations which exceeded limits, specifically for Chromium (n=34) in Bonje; and Vanadium (n=22) in Area 25. I conducted BKMR analyses on this matrix, as even at low concentrations, this method can detect metals interactions, and their effects on eGFR. Initial linear regression models showed no association between any blood metals and eGFR; however, linear models do not account for non-linear or joint effects of metals and therefore high correlations between metals in a linear model could lead to distortion of results [329].

The results from the BKMR in Bonje participants indicated that Cadmium had a significant negative effect on eGFR, which seemed to be stronger at the lower percentiles (25^{th} and 50^{th}), and when all other metals are held at their respective levels. Throughout the medical literature, Cadmium is a well-documented nephrotoxicant [330-333]. Generally, epidemics have been characterized by high dose exposure to Cadmium from water sources or diet, with resulting high concentrations measured in bio samples and tissues[33, 34]. However there is a growing body of research suggesting that chronic-low dose exposure could also be detrimental to kidney function[118, 334] which in this setting could warrant further investigation. In Area 25, there was no significant overall effect of the heavy metals' mixture on eGFR. The univariate metals overview indicated that Nickel has a significant negative effect on eGFR but only at the 25^{th} and 50^{th} percentiles. A further negative effect was observed with Chromium; however, this association was not significant. The remainder of the metals in the mixture showed a positive effect on eGFR, however none of these effects were significant. Water could not be included in the BKMR model as multiple participants were assigned to few water sources which resulted in repetitive values for heavy metals which could not be modelled efficiently. However, as there is no definitive link between water and blood concentrations with eGFR in this sub sample (and in the wider CKDu literature), biomarker data are the most appropriate to use in this instance.

In the water sources, the mean heavy metal values did not exceed permissible limits. There were however site-specific differences observed for hardness, Barium and Nickel, all of which were higher in wells in Area 25. This finding corroborates with the literature that surface

water - particularly wells - contain higher concentrations of heavy metals, which include Barium, Nickel and hardness[30, 99]. Although not fully quantified, it is thought this is due to the high electrical conductivity of uncovered standing water, which encourages metals to dissolve [78, 129, 335]. In addition, heavy metals in wells can be higher due to surface water runoff from nearby agricultural or industrial activities which can cause an increase in heavy metals concentration in comparison to other protected water sources[297].

As water and blood could not be modelled together using BKMR, I used a Pearson Correlation matrix to test for correlations between blood and water metals, followed by linear regression stratified by study zone to investigate potential associations of these matrices with eGFR. In Area 25 there was a significant negative association between blood and water Chromium concentrations and eGFR which could suggest that drinking water is a mediating factor for low eGFR in this setting. Investigation into this association would however require more detailed future investigation.

After characterization and analysis of the heavy metals data, I conducted regression analyses on all the variables in the harmonized dataset. Due to sparse data, I conducted variable profiling of all covariates in association with eGFR in linear and logistic regression models due to concerns of inflation of risk ratios but there was minimal effect across the models so I analysed eGFR on continuous scale.

I observed variance inflation factors which were close to the threshold between heat index and study zone and altitude and study zone (VIF = 7,8 respectively), so I therefore decided to model the fully adjusted model without the study zone variable as I was most interested in assessing the relationship of the environmental variables with eGFR more than the effect of a study zone in this case. In multivariate linear regression models increasing age and concentrations of blood aluminium levels were associated with a lower eGFR. Due to the cross sectional nature of this study, I am not able to decipher a causal association for this Al-eGFR relationship (i.e. if this blood concentration of from a potable water source or foodstuff), therefore this would require further investigation in future, perhaps via a longitudinal cohort study.. Furthermore, small holder farmers had a positive association with eGFR. No other associations were observed in this model, which could in part be due to the small sample size. As there was potential multicollinearity between study zone and environmental variables, it

could be suggested that variables such as heat index and altitude could be used as surrogates for study zone.

In the case of heat index modelling, the sensitivity analyses I conducted suggested that EBK was the less appropriate method to use for the ERA-5 data across Malawi, and that bilinear interpolation (restricted by study zone altitude) is likely the most appropriate model which had the lowest overall errors. When modelling the variables in the harmonized dataset using the bilinear interpolation derived heat index data, there were some associations which were lost. Most notably, the negative association between Al in the blood and eGFR, and the positive association between small holder farmers and eGFR. This could be indicative of potential exposure misclassification of heat index data in the EBK heat index models which could have produced some spurious associations with eGFR. This could however also be due to the very small sample size of this Malawian sample.

The major limitation of this study was sparse data due to a smaller sample size than anticipated, and a very low number of participants with a low eGFR. As a result, eGFR could only be modelled on a continuous scale, and risk ratios and linear mixed models were not modelled. Secondly, there were significant challenges in tracking participants to the real water source they were using due to illegality in well use. I therefore I had to assign the population to the nearest source using average walking distance information. There was therefore a high chance of water source misclassification throughout this population which could affect the reliability of the results seen in the linear and logistic regression models. The method I used to assign nearest water source to participants was conducted using the advice and guidance from multiple field workers who were local to the areas, and therefore this was the most appropriate approach to take under the circumstances. This resulted in multiple repetitive values for water heavy metals which could not be modelled in the BKMR analyses. Due to early termination of the overarching DEGREE study, I was not able to return to Malawi to conduct dry season water source collection which would have allowed for a seasonal characterization of the water sources. Finally, similar to chapter 5, there was participation bias whereby most of the respondents to the nested study were located in Bonje which could have affected the reliability of the results to the extent that the rural sample population was much better characterized and investigated in comparison to the rural population.

The main strengths of this study had been the collection of detailed lifestyle and environmental data which can be used for future studies or international comparisons. Furthermore, biomarker data to measure the “internal” concentration of heavy metals in this population was a significant strength.

This type of environmental study is the first of its kind in both Malawi and in the DEGREE research framework. The datasets and methods that I have used in this framework have proven to be useful for conducting detailed environmental analyses in suspected CKDu regions.

This framework is currently being piloted in other DEGREE sites and could have scope to be used in other areas of environment-disease research in future.

Overall, this study shows that there is little association between environmental variables and low eGFR in this sub sample. Linear regression models indicate that the key variables associated with a lower eGFR are age and increasing altitude, and that HI has a significant positive association with eGFR. Blood and water heavy metals were below permissible limits however, via BKMR analyses negative associations were observed between blood Cadmium and eGFR in rural Bonje, which could support the growing evidence that low dose Cadmium exposure on negatively impact renal function. Analysis of blood and water metals with eGFR showed a negative association between Chromium in both matrices, along with well water sources, which suggests that wells are a potential mediating factor in low eGFR in this setting – even at low doses. Due to sparse data however, these results should be interpreted with caution. This study is the first of its kind in Malawi, and across the wider DEGREE research group and the structure and methods could prove to be valuable in aiding international comparisons between DEGREE study zones in future.

7 General discussion

7.3 Summary of findings and their public health implications

CKDu is an emerging epidemic in several regions of the world and has been linked to a wide range of environmental risk factors which have generally not been investigated in detail. This thesis has described exploratory research on associations between widely hypothesised environmental risk factors and low eGFR across CKDu-endemic countries Sri Lanka and India, and suspected at-risk populations in Malawi. The motivation for this work was twofold. Firstly, I wanted to investigate whether risk factors such as temperature measured using the heat index metric, altitude, and pesticide exposure measured by vicinity to cropland as a proxy are associated with low eGFR, and whether these risk factors are common to all CKDu endemic and/or suspected regions. Secondly, I wanted to conduct a detailed environmental investigation collecting a range of variables including bio samples and potable water sources to measure heavy metals toxicity, and lifestyle and occupational information in suspected CKDu endemic Malawi as currently, studies in existing endemic regions lack this information which is a major limiting factor for environmental risk factor analyses and international comparisons. I therefore designed a nested environmental study and methodology framework to investigate these factors which I used in part for Sri Lanka and India, and in detail in the Malawian population study sample.

The specific objectives of this thesis were:

(i) To conduct a global detection of CKDu using the Global Burden of Disease dataset by applying the "passive detection" method to assess whether use of routinely collected data sources are appropriate for disease detection

(ii) Undertake exploratory environmental analyses in Sri Lanka and India where CKDu has been reported, to investigate whether environmental risk factors temperature, altitude, vicinity to agricultural land and farming practices are associated with a low eGFR

(iii) Analysing the DEGREE cross-sectional study in Malawi, which is considered at risk of CKDu, to determine the distribution and prevalence of low eGFR (<60mL/min per 1.72m^2) in an urban and rural population

(iv) Using the DEGREE cross-sectional study as a sample frame for a nested environmental study in Malawi investigating individual level environmental exposures:

a. Heat
b. Altitude
c. Heavy metals in potable water and blood samples
d. Physical symptoms of heat stress resulting from occupation.

Addressing objective iv will provide a framework to collect individual level environmental risk factors and understand if these largely hypothesised factors are associated with CKDu in these regions. This is important as it seems that the characteristics of being "hot", "low-lying" and "agricultural communities" drive the location of studies – and these might not be the determining factors.

As discussed in chapter 2, the literature has linked CKDu to key environmental risk factors. However, these factors are generally region-specific, and researchers only tended to explore specific pre-conceived hypotheses. Further major limiting factors are the variability of study designs and differing methods used to measure eGFR which have made international comparisons of eGFR and environmental risk factors particularly challenging. There has been an urgent call for standardization in measurement of eGFR in the CKDu literature, which has been addressed by the DEGREE protocol. However, the scope of this protocol does not include detailed surveillance of environmental risk factors in suspected endemic regions. This issue can be overcome (to an extent) by conducting exploratory studies across endemic areas, which can serve as the basis for future research including cohort studies, exposure assessments, and intervention studies.

There is also a valid counterargument to the implementation of costly standardized studies in resource constrained areas[2] which is to provide alternative and less costly methodological approaches to measure CKDu via the "passive detection" approach. This was an aspect I tested on the global scale using a comprehensive dataset (GBD database), as to date no studies have been conducted on this scale to detect the disease. By disentangling ICD-10 mortality codes I investigated whether the signals detected were in known endemic regions, and broadly located across the developing nations of the tropics (as hypothesised).

At the start of this study, I expected that CKDu would be associated with higher temperatures, lower altitude, and high concentrations of heavy metals, particularly from groundwater sources, that cases identified would be agricultural workers exposed to pesticides; and that by virtue of having similar geographical and environmental characteristics to CKDu endemic regions, there was a high probability that the disease would be present in Malawi. The various study components that are covered in this thesis:

(i) passive detection of "non-diabetes-related CKD",
(ii) analysis of satellite-derived environmental data with low eGFR in endemic Sri Lanka and India
(iii) cross sectional study to ascertain if CKDu is present in Malawi
(iv) nested study of individual-level environmental risk factor analysis using questionnaire data, biomarkers for heavy metals and questionnaire data

showed unexpected results.

The long-term aim of my research is for this methodological framework to be used in other DEGREE sites and to enable future global environmental comparisons.

In this chapter I provide an overview of the main findings in this thesis, followed by a discussion on each aspect of the research I conducted, beginning with the global passive detection analysis of "non-diabetes related CKD" to determine whether this method is appropriate to apply in resource constrained settings, and then discussing each of the risk factors studied, and the implications for future research.

7.4 An overview of the findings in this thesis

The use of GBD data to "passively detect" "non-diabetes related CKD" proved to be an unreliable method in detecting at-risk regions, as the rate ratio signals of the condition were not detected in same the regions of high excess mortality due to "non diabetes related CKD" was detected. Some studies in the literature have used national renal registry and/or hospital admissions data to determine the incidence[9] and geographical distribution[10] of CKDu; however the major issues highlighted in these studies were the lack of uniform reporting of the population receiving dialysis, and the inability to discriminate between those with "traditional" CKD from those with potential CKDu. Having said this, the value of initial exploratory studies must not be discounted, as they can provide useful background regional-

level information. In CKDu-endemic Sri Lanka, no associations were observed between CKDu and environmental risk factors such as heat index, altitude, or vicinity to cropland. However, chena cultivation was associated with both low eGFR and risk of eGFR<60, which could support the "heat plus" hypothesis that loss of renal function is due to high ambient temperatures exacerbating strenuous working conditions, and not temperature alone. In India, vicinity to cropland was associated with low eGFR, as was increasing altitude. The cross-sectional DEGREE study in Malawi, showed that low eGFR is not a public health issue, and that CKDu is non-existent in this population. Overall, the risk factors of increasing age and BMI were significantly associated with a low eGFR, however, these are "traditional" CKD risk factors which are likely associated with natural renal decline over the life course. The nested environmental study is the first of its kind in both Malawi and across any DEGREE study site. The results showed that overall, increasing age was associated with low eGFR, as was increasing altitude. Water heavy metals did not exceed permissible limits, however concentrations of Barium, Nickel, Cadmium, chromium and water hardness were higher in wells in Area 25. Pearson correlation matrices showed no correlation between any of the heavy metals in water and blood in Area 25. In Bonje there was a correlation between Chromium in water and blood which warrants further investigation, as drinking water source could be acting as a mediator for lower eGFR. Finally, BKMR models showed that in Bonje, Cadmium at low concentrations has a significant negative impact on eGFR. In this population, few participants reported consistently raised heart rates as a result of their occupation or daily activities. This suggests that, regardless of ambient temperatures, in this Malawian population physically strenuous work is not a factor in compromised eGFR. There were some inherent limitations to the nested study in Malawi, most notable being the very small sample size. This resulted in the need to conduct variable profiling of all the variables with eGFR to investigate independent associations before they were added into a model, and furthermore resulted in logistic regression models being very limited in their use.

7.5 Public health implications of CKDu

CKDu carries many implications for health providers and services, many of which are yet to be quantified. In high-risk communities, CKDu is being diagnosed in the late stages, when generally the individual requires renal replacement therapy (RRT). This late diagnosis is largely due to the clinically asymptomatic nature of the disease until late stages, and also a lack of CKD

surveillance at the primary, secondary, or tertiary health care level. This is due to the fact that many of these affected countries still have a substantial infectious disease burden, which is a public health priority. Furthermore, there is an inherent lack of education on both CKD and CKDu, and many clinicians are likely unaware of the current CKDu definition, and specific co-morbidities to measure when a patient presents with urgent need for RRT. The impact of CKDu is growing, and its prevalence is likely underestimated across many endemic regions. Historically, renal registries or regular surveillance systems to identify population distribution patterns and trends of CKDu have been largely heterogenous and therefore unreliable. Overall, the prevalence estimates of CKDu have been based on patient records during treatment for end-stage kidney disease or small community research surveys. It is also thought that many rural residents tend to self-treat ailments or seek ayurvedic treatment (specifically in South Asia) [336], and do not seek western medical care until symptoms are more severe, which is exacerbating the burden placed on local healthcare providers which need increasingly costly treatment options to treat those most severely affected[337]. As per DEGREE criteria, the prevalence of suspected CKDu in Sri Lanka was 5.8%. Although current direct international comparisons are difficult, this finding suggests that the CKDu burden in Sri Lanka could be second to that in Central America where overall rates of CKDu have been reported to range from 10-20%[3, 4]. In the analyses I conducted in Sri Lanka, chena cultivation was associated with the risk of having an eGFR <60 - another factor akin to that in Central America. There will likely be significant near- and long-term burden placed on the Sri Lankan health service, which is currently sparse in rural regions, and in regions where treatment is accessible, it is expensive. Rural kidney patients generally travel hours to a referral hospital to receive RRT[337], which places time and economic burdens on families. As a result, it is recognised that many forego treatment because of pressure of work, and social stigma and are likely to die prematurely [338, 339]. Proper treatment of CKDu in agricultural communities requires investment focused on structural reconfiguration and strengthening of healthcare services, and timely treatment delivery[340]. The high cost and demand for remedial treatment, and a shortage of clinicians skilled in treating CKDu present a significant and growing challenge for the public health care system in Sri Lanka.

In contrast, low eGFR in Malawi was not identified as a public health issue, and CKDu was non-existent (at least to the extent it could be investigated). Whilst this observation is reassuring

from a public health perspective, this disease cannot be completely ruled out from other African countries. As discussed in chapter 3, many nations in Africa are undergoing different stages of an epidemiological transition whereby an increase in prevalence of NCDs is being observed. Increasing CKD prevalence is most likely linked to the effects of antiretroviral therapy (ART) for HIV. However, the interplay of ART and non-communicable disease is complex highlighting social and behavioural changes which will adversely impact progress in improving health and reducing preventable mortality across Africa. An integrated healthcare plan and program delivery are required to improve access to non-communicable disease prevention and treatment. This can be achieved by establishment of surveillance sites, like The Malawi Epidemiology and Intervention Research Unit (MEIRU) which conducts ongoing long-term epidemiological studies on local urban and rural populations. Via prospective long-term surveillance and testing strategies, more targeted medical care could be provided. The most pressing limiting factor to measuring the burden of CKD in Africa however, is the lack of verification of the CKD-EPI equation in African populations[49, 174]. Currently, verification studies are underway to provide guidelines for an optimal measurement for renal function across Sub-Saharan populations[49], which will be invaluable for future renal research across African countries.

7.6 Passive detection

Overall, CKDu has been reported across developing nations which have low per capita spend on healthcare, variable diagnostic methodologies, and coding for cause of death, and largely unreliable reporting of vital statistics. Due to these issues (and others discussed in chapter 3) routinely collected data are not granular enough to confirm a strong signal of the disease, or indeed to identify at-risk populations. These deficiencies are confirmed in a global systematic review of the quality of 144 renal registries[341]. The review found that the depth and reliability of information on factors including dialysis modalities, comorbidities, and clinical outcomes were sparse to non-existent, and that even where these data were adequately collected, data accessibility is often difficult. Furthermore, limited data were available from developing nations, resulting in vast information gaps about remedial treatments and outcomes for large populations of CKD patients. In community-level CKDu studies the limitations of renal registry data have also been highlighted. Most notably these limitations were misclassification[10] and overestimation of CKDu from use of dialysis enrolment as proxy for the disease[9]. CKDu

remains a challenging disease to detect and diagnose, and certainly requires a global effort to accurately quantify its burden using the standardized methods outlined in the DEGREE protocol. That said, the question still remains on how low-income nations which are suspected as high-risk can detect this disease without having to set up future cross-sectional and cohort studies.

7.7 Heat index and altitude

The study zones across India and Sri Lanka had relatively similar maximum temperatures (31.2 and 33.5^{o}C respectively) and were situated below 400m altitude. Overall, these characteristics concur with the literature that CKDu occurs in hot, low lying regions [10, 17, 256, 257]. In Malawi, the urban and rural populations resided at higher altitudes (500-1125m), and cooler temperatures (19.3-22.2^{o}C) than those reported in other CKDu-endemic regions[14]. Despite these differences however, multivariate regression models applied across Malawi, Sri Lanka and India showed no association between increasing HI and low eGFR. This interesting observation does not support the hypothesis of heat stress nephropathy with persistent heat exposure and inadequate rehydration leading to CKD[31]. *Herath et. al.*[342] counter argued that there is sparse evidence of CKDu in workers exposed to heat in most tropical regions, supported by the finding that it is also seen in people who are not exposed to heat stress in these affected regions; their opinion is that there is inadequate evidence for heat being the initiating or main cause of CKDu. Similarly, *Ordunez et.al.*[15] in their analysis of time trends and mortality patterns for CKD in Central America, suggested that the heat hypothesis cannot fully explain the epidemic, as temperatures have not increased at a rate which could support such drastic loss of kidney function[15, 288].

The issue of interpolation models and their respective benefits must be addressed. I initially used the EBK predictive modelling method, incorporating the entire modelled dataset across all study countries (India, Sri Lanka, and Malawi) and then scaled the interpolation models down to the study zone level. I conducted a sensitivity analysis for these models across each country, by comparing the more complex EBK method with more simplistic methods inverse distance weighting, bilinear interpolation, and kriging. For these models I restricted the original heat index remodelled data to an altitude range within 100m of the study zones across each country and lowered the search radius from 10-15 neighbours to 2-4. Overall, results showed that in fact bilinear interpolation created the best, and lowest error model fit across all countries. In the case of the ERA-5 data at its current resolution and the sample sizes I was

working with, it is suggested that this more simplistic model should be used going forward – especially for local study teams who might not have a strong in GIS and analysis of satellite data.

In my analyses across this thesis, a negative association was observed between eGFR and increasing altitude in India. This finding does not concur with the literature, as generally a higher altitude (and therefore lower ambient temperatures) has been associated with a higher eGFR particularly amongst those working in physically strenuous occupations[14, 257]. Studies investigating "traditional" CKD, however, have reported similar negative associations between increasing altitude and low eGFR. A study conducted by *Meera et.al* in Costa Rica[19] found that the risk of CKD increased with altitude of residence after adjustment for socioeconomic variables. Like Sri Lanka, India and Malawi, this Costa Rican population resided between sea level and 1,200m. It was postulated that renal injury was a result of HARS (discussed in chapter 2) at the highest altitudes in the study population[19]. This is likely not the case for Sri Lanka or India because HARS is synonymous with hypertension and proteinuria which are not CKDu co-morbidities; however, hypertensive and proteinuric participants in the Malawian nested study I conducted were not excluded. The HARS hypothesis could be a plausible hypothesis for low eGFR in this population, however this would require further detailed analyses to quantify.

Most studies investigate associations between variables and eGFR using linear and logistic regression models; however these models do not account for the fact that observations in the same cluster (or study zone) are more correlated than those in another cluster [253]. To investigate the impact of environmental variables on eGFR alone I used linear mixed models to account for the above. These models assumed that the average eGFR was different between study zones, and that the effect of environmental variables on eGFR were stable across the study zones. I applied LMMs to Sri Lanka and India only, as the sample size in Malawi was too small. Comparative LMM models showed a significant negative association between eGFR and altitude, and a weak protective effect with increasing HI, particularly in Sri Lanka. To the best of my knowledge, the use of LMM models to identify associations between eGFR and environment has not previously been conducted in regions at risk of CKDu, and this could be a valuable method to use in future environmental investigations.

As discussed in chapter 1, there is a real need to move beyond ecological studies which address single environmental exposures in a population sample, and to assess exposure at the individual level. I used satellite data to assign heat index, land cover and altitude variables to

participant coordinates across all the study areas with the aim of understanding whether these purported risk factors are indeed associated with reduced eGFR and/or risk of low eGFR. From my analyses it could be suggested that not all environmental risk factors are homogenous across all high-risk CKDu countries but are perhaps more region-specific, which could help to guide future hypothesis testing, an move the research agenda towards exposure assessments at the individual level across all DEGREE sites.

The use of satellite-based data and GIS has been a valuable tool in the characterization of environmental variables across India, Sri Lanka and Malawi, as, across these countries there is still a lack of ground-based data to work with. The use of satellite data at relatively high resolutions allowed for environmental variables to be assigned to all participants in very small areas which would otherwise have no observations available.

There were however challenges which arose around obtaining satellite data with the same resolution across the three countries – specifically with regards to land cover data. In India and Sri Lanka, the highest resolution for land cover data was 300m from a validated satellite dataset. In Malawi however there was a higher resolution dataset available (20m) which was a prototype dataset at the time of analysis. The inherent limitation with the prototype was that there might have been significant error in the satellite imagery which could have resulted in misclassification of land cover classes to participants in the nested set; however, as the study zones in Malawi were very small this prototype was the most appropriate dataset for the area. At the time of conducting my analyses, the 30km resolution ERA-5 temperature data was the most recent available dataset. Since I have completed my analyses a new, higher resolution (10km) temperature dataset has been released. Had this dataset been used for my main analyses it would have altered my methods used to interpolate the HI data. I would have been able to create models at the zone level (India) and the state level (Sri Lanka and Malawi) instead of the wider country level (as the resolution of the climate data was much lower than the size of the separate study zones). I would also have conducted simpler interpolation analyses such as inverse distance weighting due to the smaller areas.

7.8 Vicinity to crop land, occupation, and self-reported symptoms

In Sri Lanka, India, and Malawi, much (if not all) of the population lives in close proximity to cropland. There were no associations observed between vicinity to a land cover class and eGFR

in Sri Lanka and Malawi, however, in India a negative association was observed with vicinity to cropland.

As there were no occupational variables in the Indian dataset, it is not possible to link this observation to a specific occupational category in this population. Using linear mixed modelling however I was able to identify the study zones with the lowest mean eGFR, and which were likely affected by vicinity to crop land, which allows for interpretation of some of the farming practices in these regions. Low eGFR was associated with vicinity to cropland in both urban and rural Andhra Pradesh and rural Haryana. Andhra Pradesh predominantly has extensive coconut and cashew cultivation[64], whereas in Haryana the predominant crops are wheat and rice[212]. In Andhra Pradesh, there are an increasing number of studies investigating the high prevalence of CKDu in relation to agricultural occupations; however to date no association has been observed between farming practices and low eGFR or CKDu in this area[64]. As mentioned in chapter 4, a DEGREE survey is currently underway in India, which will help to elucidate this relationship further as occupational data are collected.

There is only one study in the literature which has assessed the spatial relationship between cropland and CKDu. Based in El Salvador, *VanDervort et.al.* conducted an ecological analysis using hospital admissions data for suspected CKDu and estimated the percentage area of different crops across the country[10]. Results from multiple stepwise geographically-weighted regression models suggested that proxy CKDu admissions rates were associated with percentage area of crops including sugarcane, corn and coffee[10]. This finding corroborates with much of the wider Central American CKDu literature that sugarcane cultivation is a major risk factor for those at risk of, or suffering from the disease[13, 14, 29, 82, 181]. This study highlights the limitations of the ESA-CCI land cover data I used in my analyses. Despite the high resolution of the land cover classes in the dataset, the lack of spatially explicit crop type is a significant limiting factor, as the areas of highest prevalence of CKDu (Sri Lanka an India) or lower eGFR (Malawi) could have been linked to a specific crop (and associated farming practice) as opposed to the vicinity to a land cover class alone.

The Sri Lanka dataset contained additional information on variables including farming practices, of which chena cultivation posed a significant risk of having an eGFR<60. As discussed above (and in chapter 4), Chena cultivation involves the "slash and burn" of jungle or forest before cropping, and is considered to be highly strenuous [259]. In Central America similar "slash and burn" used in cultivation of sugarcane farming is implicated as a major risk

factor for CKDu [13, 29, 82, 186]. This common finding in two geographically distinct countries of strenuous type of labour suggests this could be a common risk factor for CKDu globally. Repeated volume depletion from strenuous muscle exertion has been implicated as a possible causative mechanism of CKDu[12, 14, 16, 289]. Severe volume depletion may precipitate rhabdomyolysis where breakdown of skeletal muscle releases potassium, phosphates, and myoglobin into the blood circulation resulting in severe acute kidney damage. Volume loss has been observed in sugarcane cutters and cotton workers in conditions of extreme heat stress [11, 13, 14, 82]. From these observations, it is possible that strenuous agricultural labour is a significant risk factor for CKDu. Although there was no strong association with heat index, this finding could support the "heat-plus" hypothesis[26] whereby exposure to high ambient temperatures is exacerbated by strenuous working conditions which cause CKDu, rather than heat being a direct causal factor[288, 342]. Furthermore, it could also be possible that physical exertion is exacerbated by other unmeasured confounders such as daily intake of water or high sugar drinks, frequency of rest breaks at work, or the effects of viral illness resulting from snakebite which are common in the region, and can cause renal damage [213, 343].

Across Malawi, most of the population (both urban and rural) was involved in subsistence agricultural work and the majority of crops cultivated were rice, maize, and cassava. Rice cultivation is known to be physically strenuous[344], however very few participants reported having a raised heart rate as a result of their agricultural activities. Most reported thirst and sweat, which are indicative of strenuous labour, however the extent was not possible to quantify. My observation over the 3 months in Malawi was that subsistence farming is less strenuous, with individual farmers working small plots, having more control over their working conditions, rest breaks and hours worked per day.

7.9 Water Sources

Data on drinking water sources was only available for Sri Lanka and Malawi. In chapter 4, I observed that in Sri Lanka, drinking from water sources other than groundwater had a positive effect on eGFR. As discussed in chapters 2 and 4, toxicity from drinking water is considered a central causative factor for CKDu, with studies in Sri Lanka recording that regions with the highest CKDu prevalence used wells as their primary water source[30, 129, 258]. In response to the increasing CKDu reports in the country, Sri Lanka is focusing on research into the remediating effects on water purity of reverse osmosis (RO) and bottled water in CKDu

endemic regions[260, 261]. A study conducted by *Athapattu et.al.,* showed that treatment of rejected wetland water with RO considerably reduced potentially harmful parameters including pH, heavy metals and hardness, and met the permissible drinking water limits for the region post-treatment[260]. The establishment of RO is proposed as a short-term measure to combat lack of access to clean, non-toxic water as it is costly to implement, and therefore costly for many local residents[261]. Although groundwater analyses in Sri Lanka have not provided conclusive evidence on the causative nature of water sources in CKDu [130, 213, 215],there remains wide support for toxicity of well water as an important factor in CKDu in endemic regions.

In chapter 6, I observed that water sources from wells in Malawi were associated with a low eGFR in univariate modelling. There is very little information on quality of water by source in Malawi; those studies which have been conducted have largely focussed on bacterial coliforms, as cholera is a major public health issue across the country[345, 346]. In a study conducted by *Mkwate et.al*, wells had a higher concentration of coliforms in comparison to boreholes, and were deemed unreliable water sources[347]. A further study[348] into microbiological quality of water from shallow wells in rural Malawi confirmed that these sources fall below acceptable standards and called for sustainable cost effective technologies to treat groundwater for rural communities dependent on these sources. Although bacteriology studies cannot provide information on heavy metals, they provide a potentially useful indicator for poor quality of well and borehole water.

7.10 Heavy metals in potable water

In Malawi, I found that mean heavy metal concentrations in potable water did not exceed permissible limits. This finding is corroborated by some of the Sri Lankan water quality studies, which report low concentrations of heavy metals in endemic regions [61, 349]. There were however site-specific differences observed in my study for hardness, Barium and Nickel, all of which were higher in wells in Area 25. This finding corroborates with the literature that surface water - particularly surface wells - contain higher concentrations of heavy metals, which include Barium, Nickel and hardness[30, 99]. Although not fully quantified, this is due to the high electrical conductivity of uncovered standing water, which encourages metals to dissolve thus increasing toxicity [78, 129, 335]. In addition, heavy metals levels in surface wells can be higher due to surface water runoff with high heavy metal content from nearby agriculture or industry [297]. Common water sources in Malawi were wells and boreholes, and most of the population

reported a high daily intake of water from these sources. As discussed above, wells were significantly associated with a low eGFR, however, no associations were observed with daily water intake (~ >2 cups per day) and water treatment methods. This suggests that the water quality in Malawi, most likely has little impact on eGFR and as a factor in Malawian CKD.

7.11 Heavy metals biomarkers

In whole blood samples, the mean heavy metals concentrations did not exceed permissible limits, however, they were marginally higher in Bonje. The biomarkers were analysed using the novel method BKMR, which offers distinct advantages over linear regression models as it allows for potential and synergistic effects amongst mixture components, and simultaneously provides measures of variable importance. Overall, heavy metals concentrations did not exceed reference ranges. However, in rural Bonje Cadmium had a significant negative effect on eGFR, which was stronger at the lower percentiles, while all other metals are held at their respective levels. This finding is similar to those of *Sanders et.al.* who found that Cadmium was a significant driver of renal damage when assessing co-exposures of blood metals (Lead, Cadmium, Mercury) on kidney function in adolescents[331].

In the medical literature, Cadmium is a well-documented nephrotoxicant [98, 118, 305], affecting renal function by damage to the proximal tubules [111]. As discussed in chapter 6, epidemics of CKD have been characterized by high dose exposure to Cadmium from water sources or diet, with resulting high concentrations measured in bio samples and tissues[33, 34]. In the CKDu literature, studies investigating heavy metals biomarkers have shown mixed results. There is however a growing body of research suggesting that chronic low dose exposure could also be detrimental to kidney function[118, 334]. The findings from the BKMR analyses support this hypothesis, as Cadmium concentrations did not exceed permissible limits in blood or water but did have a significant negative association with eGFR.

There is debate as to which matrix is optimal for measuring heavy metals biomarkers. In the literature, most studies have used urine, and fewer have used blood, fingernail, and hair samples[20, 66, 350, 351]. Across these studies, differing concentrations of metals were observed in each matrix. In a study conducted by the WHO-CKDu Group[66], random urine, nail and hair samples were taken from participants with and without CKDu in Sri Lanka. Amongst the CKDu patients, a higher level of Cadmium and Arsenic was observed in nails and hair, but a lower concentration of urinary Arsenic and Cadmium in comparison to the control group. This

interesting finding corroborated with other studies which found that increased urinary Cadmium was associated with an increased eGFR[352]. In a Sri Lankan case-control study, *Nanayakkara et.al.* reported urinary Cadmium excretion to be lower in CKD*u* patients compared to controls. This observation was consistent with the findings of Bandara *et.al.*, who hypothesised that an inability to express the urinary chelating protein metallothionine could lead to a reduced urinary Cadmium concentration and increased tubular damage[25]. Differing results could be explained, in part, by heterogeneity of study design, control selection, and diverse means of assaying Cadmium. This factor could also be due to reverse causality whereby a low eGFR could result in a decreased excretion of Cadmium from the urine and therefore increased levels circulating in the blood[353].

Most of the heavy metals studies in the literature are cross-sectional and therefore single measurements of environmental chemicals may not represent true exposure profiles that can be compared with health outcomes.

7.12 The issue of causality in CKDu research

As discussed throughout sections of this thesis, the issue of causality in CKDu is complex, and there are yet to be any definitive associations made among environmental risk factors. It has been said that causal inference requires a pluralistic approach, with the integration of a wide range of methods, study types and evidence in order to elucidate mechanisms in environmental exposures and population health[354-356]. Pearce et.al.[357] suggest that causation in environmental epidemiology should include: (i) use of 'traditional' methods; (ii) extensions of traditional approaches; and (iii) triangulation of evidence[357]. Although causality could not be established in this research, an attempt can be made to investigate how causal associations could be investigated based on the results I present in this thesis, and help to make suggestions for the direction of future CKDu research.

Traditional methods

Studies assessing the geographical and international differences of disease prevalence are particularly useful, both in creating new ideas, and as an investigation to check existing explanations for the potential causes of a disease. Furthermore, findings can be more convincing when conducting cross-context comparisons[358]. In the case of CKDu, the disease is

currently being investigated across multiple study zones globally, results of which are now playing a crucial role in establishing countries which have a high prevalence of the disease. As discussed throughout the thesis, the DEGREE framework does not aim to diagnose the disease as it uses a cross-sectional methodology; therefore, the issue of causality cannot be addressed using this framework alone. When used in tandem with the environmental framework across suspected high-risk regions however, one can compare prevalence and environmental exposures together. As a result of this collection of data from different countries, evidence from these studies (and sources of bias) can be gathered and guide longitudinal studies in countries which have a high prevalence of CKDu, thus allowing for a more targeted study on diagnosing CKDu and determining which of the environmental risk factors (or collection) could be causal to CKDu.

Extensions of traditional approaches

An extension of the "traditional" approach could be to use Mendelian randomization (MR) that use gene-environmental interactions to investigate causal effects in future CKDu studies once the key environmental risk factors for the disease have been established across affected study zones[359]. To do this, the approaches to future cohort study designs would have to evolve. In order to understand if there is a genetic underpinning to CKDu in a given population, there must be the necessary elements in place to detect this, such as development of affordable genotyping for different populations, and promotion of comprehensive single nucleotide polymorphism array genotyping for high prevalence CKDu areas to identify a potential major gene effect in populations exposed to a specific environment[360]. The development of standardized data collection tools outlined in the environmental framework is an important addition to the above to refine the identification of global and local environmental risk factors and to develop additional renal endophenotypes in affected regions[360].

Triangulation of evidence

Triangulation - the integration of evidence from several different epidemiological approaches that have differing and unrelated sources of bias – could be used to improve causal inference in aetiological epidemiology. By integrating epidemiological, animal, and mechanistic evidence to infer causality for various potential environmental exposures. In chapter 2 triangulation was described in the case of Balkan Endemic Nephropathy, for which a wide

variety of evidence (epidemiological, genetic, toxicological) was required before it was concluded that the likely cause was dietary exposure to aristolochic acid, which was a contaminant of wheat in endemic regions across the Balkan Peninsula. The development of my environmental framework across global DEGREE zones is be a first step in creating a database of integrated data which includes biomarker, descriptive and observational data which acknowledges the key sources of bias, which will allow for better causal inference.

7.13 Implications of CKDu at the individual level

CKDu is a clinically silent disease primarily affecting economically active populations and causing the premature death of young people. The impacts of this disease are major, and place a significant socioeconomic burden on communities, and health services. Those who are diagnosed with CKDu are generally rendered unable to work and suffer financial losses and reduced ability to support their families. An indirect effect is that young adults in affected family groups will leave education early to financially support their family; this in turn and in the long-term could result in cycles of health-related poverty promoting disease across future generations.

7.14 Academic research implications

CKDu research and detection methods have developed rapidly over the past two decades, and promising novel approaches and methods to elucidate potential aetiologies of the disease are on the horizon. Clinically there is an urgent need for at-risk regions to adopt the standardized detection methods as outlined by DEGREE to allow for international comparisons. In my opinion there is also an urgent need for these communities to also adopt standardized environmental surveillance and data collection methods. The environmental risk factor studies in the literature lack a multifactorial approach and adopt very different study designs when measuring exposures. As a result, it is difficult to compare these risk factors effectively across potentially endemic regions. The findings from this thesis indicate that environmental risk factors for low eGFR and CKDu are multifactorial, and not uniform across endemic regions. It is also clear that more detailed, longitudinal follow-up studies are required in potentially endemic regions in order to effectively quantify toxicity from external sources. A research network which collects and measures environmental variables across endemic regions feeding into an "environment-wide association study" (EWAS)[361], will be invaluable in analysis of risk factors in CKDu phenotypes. To achieve this there must be standardized methods and

protocols across all sites, and most importantly, there must be open channels and access to data for global research of potential environmental risk factors. Going forward, there is a need to strengthen global environmental toxicology and environmental networks by creating and supporting transparent collection and exchange of information.

Table 61. Time and financial cost for setting up and rolling out the environmental module framework in other DEGREE sites

Variable	Duration to complete	Financial cost*
Questionnaire creation/adaptation	1-2 months for research and local adaptations	
Training field staff	2-3 hours	Included in field workers allotted pre-field work daily meeting
Conducting reconnaissance trips in study area of interest	1-2 days (depending on size and number of study zones)	Fuel price (Malawi ~ £150 GBP)
Questionnaire collection	4-12 months (dependent on sample size, time frame of study and number of study zones)	Extra 30 minutes per day of field workers time and according hourly rate (Malawi ~ £3,000 GBP)
Water sample collection	2-6 days depending on sample frame area and size	Water board scientist to provide apparatus an expertise: £400 GBP
Water sample analyses	Approximately 4 weeks	Dependent on laboratory sample charge rate Nottingham charge: ~ £2500
Satellite imagery collection	1-2 days	Free resources
Training for statistical methods BKMR, interpolation, regression	2-3 weeks	Hourly rate for teacher

*Financial costs shown are estimations relevant to the cost of set up, data collection and laboratory analyses for the Malawian sample

7.15 Key further research questions

Having researched the literature and conducted my own environmental and field research on CKDu across different regions, I can identify areas to benefit from further research. These research questions and future work ideas are:

- Firstly, and perhaps most pressing, is the need to validate the CKD-EPI equation in populations around the world to avoid underestimation of low eGFR prevalence, specifically in black populations.

- The studies conducted in this thesis were all cross sectional and therefore temporal trends of eGFR, and environmental risk factors could not be quantified. It is therefore necessary to conduct prospective, longitudinal, cohort studies with multiple repetitive sampling in DEGREE-identified at-risk communities which could help decipher causal relationships between CKDu and various potential aetiologies, and target future surveillance and treatment plans
- In the nested study, water sources and bio samples could not be measured for pesticides. It would be valuable to analyse samples for these markers and assess them relative to their contributions to eGFR, as well as assess interactions between the various markers in the blood using BKMR analyses.
- As discussed throughout the thesis, most existing studies have not considered a multifactorial approach to environmental exposures in relation to CKDu. It would be useful to set up a global consortium of researchers collecting and analyzing environmental variables as part of an EWAS study across DEGREE study sites.
- Bacterial coliforms in potable water have not been linked to low eGFR or CKDu, however this aspect would be valuable to investigate as this could capture important biomarkers for infection which could have long term effects on renal function.
- I could not use personal temperature and heart rate monitors to measure heat exposure in the Malawi population, therefore I measure this using the heat index which is the closest measure to a "feels like" temperature. In at-risk regions it would be invaluable to use personal monitors to measure the effect of heat stress over time.
- There is increasing research into algorithms to create spatially explicit crop maps over time. This type of work would be valuable at the community-level in CKDu-affected regions as it would allow for a more precise estimation of vicinity to different crop types and farming practices.
- Finally, the study zones in Sri Lanka were entirely rural and selected based on a priori hypotheses of high CKDu prevalence. To avoid the "streetlight effect" and reduce selection bias, it would be optimal to compare this rural region to an urban and /or mixed land region to understand if there is an urban/rural difference across all future DEGREE study sites.

7.16 Conclusions

In conclusion, my thesis work found that environmental risk factors for CKDu are multifactorial and not uniform to all endemic regions. The findings challenge the hypothesis that CKDu routinely occurs in “hot”, “low-lying” and “agricultural” communities. In Malawi low eGFR was not a public health concern and CKDu was non-existent, and there were no associations found with the key risk factors heat index, land cover or farming practices. In potable water samples, heavy metals did not exceed permissible limits, however there was a negative association between drinking well water and low eGFR. BKMR analyses of heavy metals biomarkers showed that Cadmium was associated with low eGFR at low concentrations. This factor requires more detailed future analyses, preferably measuring toxins in multiple matrices in a longitudinal study. Similarly, in India, there was an association between low eGFR and increasing altitude and vicinity to cropland, but not heat index.

In Sri Lanka, where CKDu is endemic, there were associations between low eGFR, altitude and chena cultivation, which could suggest that heat stress via high impact stress working conditions could be a common risk factors across endemic regions. Furthermore, drinking from water courses other than wells had a protective effect.

The question remains however as to whether work-related heat stress in higher temperatures is related to the increasing incidence of the disease in tropical areas, or whether these factors are secondary to other environmental exposures, such as heavy metals. As environmental risk factors are a major aspect of the CKDu literature, and thus influence the choice of study zones, it is vital to focus on the nested environmental study with standardized, reproducible methodologies alongside the core DEGREE survey.

Bibliography

1. Caplin B, Jakobsson K, Glaser J, Nitsch D, Jha V, Singh A, et al. International Collaboration for the Epidemiology of eGFR in Low and Middle Income Populations - Rationale and core protocol for the Disadvantaged Populations eGFR Epidemiology Study (DEGREE). BMC nephrology. 2017;18(1):1.
2. Caplin B, Yang CW, Anand S, Levin A, Madero M, Saran R, et al. The International Society of Nephrology's International Consortium of Collaborators on Chronic Kidney Disease of Unknown Etiology: report of the working group on approaches to population-level detection strategies and recommendations for a minimum dataset. Kidney international. 2019;95(1):4-10.
3. Lebov JF, Valladares E, Pena R, Pena EM, Sanoff SL, Cisneros EC, et al. A population-based study of prevalence and risk factors of chronic kidney disease in Leon, Nicaragua. Canadian Journal of Kidney Health and Disease. 2015;2 (6).
4. Torres C, Aragón A, González M, López I, Jakobsson K, Elinder CG, et al. Decreased kidney function of unknown cause in Nicaragua: a community-based survey. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2010;55(3):485-96.
5. O'Callaghan-Gordo C, Shivashankar R, Anand S, Ghosh S, Glaser J, Gupta R, et al. Prevalence of and risk factors for chronic kidney disease of unknown aetiology in India: secondary data analysis of three population-based cross-sectional studies. BMJ open. 2019;9(3):e023353.
6. Ruwanpathirana T, Senanayake S, Gunawardana N, Munasinghe A, Ginige S, Gamage D, et al. Prevalence and risk factors for impaired kidney function in the district of Anuradhapura, Sri Lanka: a cross-sectional population-representative survey in those at risk of chronic kidney disease of unknown aetiology. BMC Public Health. 2019;19(1):763.
7. Excellence NIfHaC. Chronic kidney disease. Early identification and management of chronic kidney disease in adults in primary and secondary care. London, UK: NICE; 2014.
8. González-Quiroz M, Pearce N, Caplin B, Nitsch D. What do epidemiological studies tell us about chronic kidney disease of undetermined cause in Meso-America? A systematic review and meta-analysis. Clinical kidney journal. 2018;11(4):496-506.
9. Laux TS, Barnoya J, Guerrero DR, Rothstein M. Dialysis enrollment patterns in Guatemala: evidence of the chronic kidney disease of non-traditional causes epidemic in Mesoamerica. BMC nephrology. 2015;16:54.
10. VanDervort DR, Lopez DL, Orantes CM, Rodriguez DS. Spatial distribution of unspecified chronic kidney disease in El Salvador by crop area cultivated and ambient temperature. MEDICC review. 2014;16(2):31-8.
11. Wesseling C, Aragon A, Gonzalez M, Weiss I, Glaser J, Rivard CJ, et al. Heat stress, hydration and uric acid: A cross-sectional study in workers of three occupations in a hotspot of Mesoamerican nephropathy in Nicaragua. BMJ open. 2016;6 (12) (no pagination)(e011034).
12. Roncal-Jimenez CA, Milagres T, Andres-Hernando A, Kuwabara M, Jensen T, Song Z, et al. Effects of exogenous desmopressin on a model of heat stress nephropathy in mice. American Journal of Physiology - Renal Physiology. 2017;312(3):F418-F26.
13. García-Trabanino R, Jarquín E, Wesseling C, Johnson RJ, González-Quiroz M, Weiss I, et al. Heat stress, dehydration, and kidney function in sugarcane cutters in El Salvador - A cross-shift study of workers at risk of Mesoamerican nephropathy. Environmental research. 2015;142:746-55.

14. Peraza S, Wesseling C, Aragon A, Leiva R, Garcia-Trabanino RA, Torres C, et al. Decreased kidney function among agricultural workers in El Salvador. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2012;59(4):531-40.
15. Ordunez P, Nieto FJ, Martinez R, Soliz P, Giraldo GP, Mott SA, et al. Chronic kidney disease mortality trends in selected Central America countries, 1997–2013: clues to an epidemic of chronic interstitial nephritis of agricultural communities. Journal of Epidemiology and Community Health. 2018;72(4):280.
16. Jayasekara KB, Kulasooriya PN, Wijayasiri KN, Rajapakse ED, Dulshika DS, Bandara P, et al. Relevance of heat stress and dehydration to chronic kidney disease (CKDu) in Sri Lanka. Preventive Medicine Reports. 2019;15:100928.
17. Nerbass FB, Pecoits-Filho R, Clark WF, Sontrop JM, McIntyre CW, Moist L. Occupational Heat Stress and Kidney Health: From Farms to Factories. Kidney international reports. 2017;2(6):998-1008.
18. Ghahramani N, Ahmed F, Al-Laham A, Lengerich EJ. The epidemiological association of altitude with chronic kidney disease: Evidence of protective effect. Nephrology (Carlton, Vic). 2011;16(2):219-24.
19. Harhay MN, Harhay MO, Coto-Yglesias F, Rosero Bixby L. Altitude and regional gradients in chronic kidney disease prevalence in Costa Rica: Data from the Costa Rican Longevity and Healthy Aging Study. Tropical Medicine and International Health. 2016;21(1):41-51.
20. Ananda Jayalal TB, Jayaruwan Bandara TWMA, Mahawithanage STC, Wansapala MAJ, Galappaththi SPL. A quantitative analysis of chronic exposure of selected heavy metals in a model diet in a CKD hotspot in Sri Lanka. BMC nephrology. 2019;20(1):208.
21. Jayasumana C, Gunatilake S, Siribaddana S. Simultaneous exposure to multiple heavy metals and glyphosate may contribute to Sri Lankan agricultural nephropathy. BMC nephrology. 2015;16 (1) (no pagination)(103).
22. Gunatilake S, Seneff S, Orlando L. Glyphosate's Synergistic Toxicity in Combination with Other Factors as a Cause of Chronic Kidney Disease of Unknown Origin. International journal of environmental research and public health. 2019;16(15):2734.
23. Jayatilake N, Mendis S, Maheepala P, Mehta FR. Chronic kidney disease of uncertain aetiology: Prevalence and causative factors in a developing country. BMC nephrology. 2013;14 (1) (no pagination)(180).
24. Levine KE, Redmon JH, Elledge MF, Wanigasuriya KP, Smith K, Munoz B, et al. Quest to identify geochemical risk factors associated with chronic kidney disease of unknown etiology (CKDu) in an endemic region of Sri Lanka—a multimedia laboratory analysis of biological, food, and environmental samples. Environmental Monitoring and Assessment. 2016;188(10):548.
25. Bandara JM, Senevirathna DM, Dasanayake DM, Herath V, Bandara JM, Abeysekara T, et al. Chronic renal failure among farm families in cascade irrigation systems in Sri Lanka associated with elevated dietary cadmium levels in rice and freshwater fish (Tilapia). Environmental geochemistry and health. 2008;30(5):465-78.
26. Valcke M, Levasseur M-E, Soares da Silva A, Wesseling C. Pesticide exposures and chronic kidney disease of unknown etiology: an epidemiologic review. Environmental Health. 2017;16(1):49.
27. Herrera Valdes R, Orantes CM, Almaguer Lopez M, Lopez Marin L, Arevalo PA, Smith Gonzalez MJ, et al. Clinical characteristics of chronic kidney disease of non-traditional causes

in women of agricultural communities in El Salvador. Clinical Nephrology. 2015;83(7 Suppl 1):56-63.
28. Wijkström J, González-Quiroz M, Hernandez M, Trujillo Z, Hultenby K, Ring A, et al. Renal Morphology, Clinical Findings, and Progression Rate in Mesoamerican Nephropathy. American Journal of Kidney Diseases. 2016.
29. Wesseling C, Aragón A, González M, Weiss I, Glaser J, Bobadilla NA, et al. Kidney function in sugarcane cutters in Nicaragua - A longitudinal study of workers at risk of Mesoamerican nephropathy. Environmental research. 2016;147:125-32.
30. Jayasumana C, Paranagama P, Agampodi S, Wijewardane C, Gunatilake S, Siribaddana S. Drinking well water and occupational exposure to Herbicides is associated with chronic kidney disease, in Padavi-Sripura, Sri Lanka. Environmental health : a global access science source. 2015;14:6.
31. Glaser J, Lemery J, Rajagopalan B, Diaz HF, Garcia-Trabanino R, Taduri G, et al. Climate Change and the Emergent Epidemic of CKD from Heat Stress in Rural Communities: The Case for Heat Stress Nephropathy. Clinical journal of the American Society of Nephrology : CJASN. 2016;11(8):1472-83.
32. Nakanga WP, Prynn JE, Banda L, Kalyesubula R, Tomlinson LA, Nyirenda M, et al. Prevalence of impaired renal function among rural and urban populations: findings of a cross-sectional study in Malawi. Wellcome Open Res. 2019;4:92-.
33. Gifford FJ, Gifford RM, Eddleston M, Dhaun N. Endemic Nephropathy Around the World. Kidney International Reports. 2017;2(2):282-92.
34. Nishijo M, Nakagawa H, Suwazono Y, Nogawa K, Kido T. Causes of death in patients with Itai-itai disease suffering from severe chronic cadmium poisoning: a nested case–control analysis of a follow-up study in Japan. BMJ open. 2017;7(7):e015694.
35. Stoev SD. Balkan Endemic Nephropathy - Still continuing enigma, risk assessment and underestimated hazard of joint mycotoxin exposure of animals or humans. Chemico-Biological Interactions. 2017;261:63-79.
36. Petkova-Bocharova T, Castegnaro M, Michelon J, Maru V. Ochratoxin A and other mycotoxins in cereals from an area of Balkan endemic nephropathy and urinary tract tumours in Bulgaria. IARC scientific publications. 1991(115):83-7.
37. Han Y, Shou D, Wen L, Shi J, Ding J, Gong P, et al. Interplay between chronic kidney disease (CKD) and upper tract urothelial carcinomas (UUC): Foe or friend? Oncotarget. 2016;7(33):53951-8.
38. Bellomo R, Kellum JA, Ronco C. Acute kidney injury. The Lancet. 2012;380(9843):756-66.
39. Wadei HM, Textor SC. The role of the kidney in regulating arterial blood pressure. Nature reviews Nephrology. 2012;8(10):602-9.
40. Jelkmann W. Regulation of erythropoietin production. J Physiol. 2011;589(Pt 6):1251-8.
41. Koeppen BM, Stanton BA. 2 - Structure and Function of the Kidneys. In: Koeppen BM, Stanton BA, editors. Renal Physiology (Fifth Edition). Philadelphia: Mosby; 2013. p. 15-26.
42. Pollak MR, Quaggin SE, Hoenig MP, Dworkin LD. The glomerulus: the sphere of influence. Clinical journal of the American Society of Nephrology : CJASN. 2014;9(8):1461-9.
43. Koeppen BM, Stanton BA. 3 - Glomerular Filtration and Renal Blood Flow. In: Koeppen BM, Stanton BA, editors. Renal Physiology (Fifth Edition). Philadelphia: Mosby; 2013. p. 27-43.

44. Yu JE, Lin RY. Chapter 47 - Protein-Losing Conditions. In: Sullivan KE, Stiehm ER, editors. Stiehm's Immune Deficiencies. Amsterdam: Academic Press; 2014. p. 863-73.
45. Robertson GL, Norgaard JP. Renal regulation of urine volume: potential implications for nocturia. BJU International. 2002;90(s3):7-10.
46. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, et al. Estimating Glomerular Filtration Rate from Serum Creatinine and Cystatin C. New England Journal of Medicine. 2012;367(1):20-9.
47. Davies DF, Shock NW. Age changes in glomerular filtration rate, effective renal plasma flow, and tubular excretory capacity in adult males. The Journal of clinical investigation. 1950;29(5):496-507.
48. Botev R, Mallié J-P, Couchoud C, Schück O, Fauvel J-P, Wetzels JFM, et al. Estimating glomerular filtration rate: Cockcroft-Gault and Modification of Diet in Renal Disease formulas compared to renal inulin clearance. Clinical journal of the American Society of Nephrology : CJASN. 2009;4(5):899-906.
49. Kalyesubula R, Fabian J, Nakanga W, Newton R, Ssebunnya B, Prynn J, et al. How to estimate glomerular filtration rate in sub-Saharan Africa: design and methods of the African Research into Kidney Diseases (ARK) study. BMC nephrology. 2020;21(1):20.
50. Levin A, Stevens, PE, Bilous, RW, Coresh, J, De Francisco, ALM, De Jong, PE, Griffith, KE, Hemmelgarn, BR, Iseki, K, Lamb, EJ, Levey, AS, Riella, MC, Shlipak, MG, Wang, H, White, CT & Winearls, CG. Kidney disease: Improving global outcomes (KDIGO) CKD work group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney International Supplements. 2013;3:1-150.
51. Ibrahim F, Hamzah L, Jones R, Nitsch D, Sabin C, Post FA. Comparison of CKD-EPI and MDRD to estimate baseline renal function in HIV-positive patients. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2012;27(6):2291-7.
52. Levey AS, de Jong PE, Coresh J, El Nahas M, Astor BC. The definition, classification, and prognosis of chronic kidney disease: a KDIGO controversies conference report. Kidney international. 2011;80.
53. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Annals of internal medicine. 2009;150(9):604-12.
54. NICE NIfHaCE. Chronic kidney disease in adults: assessment and management: Clinical guideline [CG182]. UK: NICE; 2015.
55. Gansevoort R, Correa-Rotter R, Hemmelgarn B, Jafar TH, Heerspink H, Mann J, et al. Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. The Lancet. 2013;382(9889):339-52.
56. Fishbane S, Hazzan AD, Halinski C, Mathew AT. Challenges and opportunities in late-stage chronic kidney disease. Clinical kidney journal. 2015;8(1):54-60.
57. Atkins RC. The epidemiology of chronic kidney disease. Kidney international. 2005;67:S14-S8.
58. Xie Y, Bowe B, Mokdad AH, Xian H, Yan Y, Li T, et al. Analysis of the Global Burden of Disease study highlights the global, regional, and national trends of chronic kidney disease epidemiology from 1990 to 2016. Kidney international. 2018;94(3):567-81.
59. Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a

systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151-210.
60. Bikbov B, Purcell CA, Levey AS, Smith M, Abdoli A, Abebe M, et al. Global, regional, and national burden of chronic kidney disease, a systematic analysis for the Global Burden of Disease Study 2017. The Lancet. 2020;395(10225):709-33.
61. Paranagama DGA, Bhuiyan MA, Jayasuriya N. Factors associated with Chronic Kidney Disease of unknown aetiology (CKDu) in North Central Province of Sri Lanka: a comparative analysis of drinking water samples. Applied Water Science. 2018;8(6):151.
62. Wijetunge S, Ratnatunga NVI, Abeysekera TDJ, Wazil AWM, Selvarajah M. Endemic chronic kidney disease of unknown etiology in Sri Lanka: Correlation of pathology with clinical stages. Indian journal of nephrology. 2015;25(5):274-80.
63. De Silva PMCS, Mohammed Abdul KS, Eakanayake EMDV, Jayasinghe SS, Jayasumana C, Asanthi HB, et al. Urinary Biomarkers KIM-1 and NGAL for Detection of Chronic Kidney Disease of Uncertain Etiology (CKDu) among Agricultural Communities in Sri Lanka. PLoS neglected tropical diseases. 2016;10 (9) (no pagination)(e0004979).
64. Tatapudi RR, Rentala S, Gullipalli P, Komarraju AL, Singh AK, Tatapudi VS, et al. High Prevalence of CKD of Unknown Etiology in Uddanam, India. Kidney International Reports. 2019;4(3):380-9.
65. Huuskonen MS, Tossavainen A, Higashi T, Okubo T. Ecological Relationship between Mesothelioma Incidence/Mortality and Asbestos Consumption in Ten Western Countries and Japan. Journal of occupational health. 1999;41:8-11.
66. (WHO) WHO. Chronic Kidney Disease of Unknown Aetiology (CKDu); A New Threat to Health. Colombo, Sri Lanka; 2008.
67. Athuraliya N, Abeysekera T, Amerasinghe P, Kumarasiri R, Bandara P, Karunaratne U, et al. Uncertain etiologies of proteinuric-chronic kidney disease in rural Sri Lanka. Kidney international. 2011;80(11):1212-21.
68. Battaglia M, Atkinson MA. The streetlight effect in type 1 diabetes. Diabetes. 2015;64(4):1081-90.
69. Orantes-Navarro CM, Herrera-Valdes R, Almaguer-Lopez M, Lopez-Marin L, Vela-Parada XF, Hernandez-Cuchillas M, et al. Toward a Comprehensive Hypothesis of Chronic Interstitial Nephritis in Agricultural Communities. Advances in Chronic Kidney Disease. 2017;24(2):101-6.
70. Orantes Navarro CM, Herrera Valdés R, Almaguer López M, Calero DJ, de Morales JF, Ascencio Alvarado NP, et al. Epidemiological characteristics of chronic kidney disease of non-traditional causes in women of agricultural communities of El Salvador. Clinical Nephrology. 2015;83:S24-S31.
71. Raines N, Gonzalez M, Wyatt C, Kurzrok M, Pool C, Lemma T, et al. Risk factors for reduced glomerular filtration rate in a Nicaraguan community affected by Mesoamerican nephropathy. MEDICC review. 2014;16(2):16-22.
72. Orantes CM, Herrera R, Almaguer M, Brizuela EG, Hernández CE, Bayarre H, et al. Chronic kidney disease and associated risk factors in the Bajo Lempa region of El Salvador: Nefrolempa study, 2009. MEDICC review. 2011;13(4):14-22.
73. Rajapakse S, Shivanthan MC, Selvarajah M. Chronic kidney disease of unknown etiology in Sri Lanka. International journal of occupational and environmental health. 2016;22(3):259-64.
74. Ramírez-Rubio OA, Juan José ; Kaufman, James S. ; Weiner, Daniel E. ; Parikh, Chirag R. ; Khan, Usman ; McClean, Michael D. ; Laws, Rebecca L. ; López-Pilarte, Damaris ; Friedman,

David J. ; Kupferman, Joseph ; Brooks, Daniel R. Urine biomarkers of kidney injury among adolescents in Nicaragua, a region affected by an epidemic of chronic kidney disease of unknown aetiology. Nephrology Dialysis Transplantation. 2016;31(3).
75. Wesseling C, Crowe J, Hogstedt C, Jakobsson K, Lucas R, Wegman DH, et al. Resolving the enigma of the mesoamerican nephropathy: a research workshop summary. American Journal of Kidney Diseases. 2014;63(3):396-404.
76. Crowe J, Wesseling C, Solano BR, Umaña MP, Ramírez AR, Kjellstrom T, et al. Heat Exposure in Sugarcane Harvesters in Costa Rica. AMERICAN JOURNAL OF INDUSTRIAL MEDICINE. 2013;56.
77. Mejía R, Quinteros, E. , López, A. , Ribó, A. , Cedillos, H. , Orantes, C. , Valladares, E. and López, D. Pesticide-Handling Practices in Agriculture in El Salvador: An Example from 42 Patient Farmers with Chronic Kidney Disease in the Bajo Lempa Region. Occupational Diseases and Environmental Medicine. 2014;2:56-70.
78. Cooray T, Wei Y, Zhong H, Zheng L, Weragoda SK, Weerasooriya AR. Assessment of Groundwater Quality in CKDu Affected Areas of Sri Lanka: Implications for Drinking Water Treatment. International journal of environmental research and public health. 2019;16(10).
79. Wimalawansa SJ. The role of ions, heavy metals, fluoride, and agrochemicals: critical evaluation of potential aetiological factors of chronic kidney disease of multifactorial origin (CKDmfo/CKDu) and recommendations for its eradication. Environmental geochemistry and health. 2016;38(3):639-78.
80. Trabanino RG, Aguilar R, Silva CR, Mercado MO, Merino RL. End-stage renal disease among patients in a referral hospital in El Salvador. Revista panamericana de salud publica = Pan American journal of public health. 2002;12(3):202-6.
81. Jimenez CAR, Ishimoto T, Lanaspa MA, Rivard CJ, Nakagawa T, Ejaz AA, et al. Fructokinase activity mediates dehydration-induced renal injury. Kidney international. 2014;86(2):294-302.
82. Laws RL, Brooks DR, Amador JJ, Weiner DE, Kaufman JS, Ramírez-Rubio O, et al. Biomarkers of Kidney Injury among Nicaraguan Sugarcane Workers. American Journal of Kidney Diseases. 2016;67(2):209-17.
83. USDL USDoL. About the heat index USA: USDoL; 2020 [Available from: https://www.osha.gov/SLTC/heatillness/heat_index/about.html.
84. Bowen-Horton S. They Leave Their Kidneys in the Fields: Illness, Injury, and Illegality among U.S. Farmworkers. Anthropology CSiP, editor. California, USA: University of California Press; 2016.
85. Jayasumana C OC, Herrera R, Almaguer M, Lopez L, Silva L.C, Ordunez P, Siribaddana S, Gunatilake S, De Broe M.E. Chronic interstitial nephritis in agricultural communities: a worldwide epidemic with social, occupational and environmental determinants. Nephrol Dial Transplant. 2017;32:234-41.
86. FAO FaAOotUN. FAOSTAT Crops USA: FAOSTAT; 2017 [Available from: http://www.fao.org/faostat/en/#data/QC/visualize.
87. Dorward A CE. The Malawi Agricultural Input Subsidy Programme: 2005-6 to 2008-91. International Journal of Agricultural Sustainability. 2011;9(1).
88. Gonick HC. Nephrotoxicity of cadmium & lead. Indian J Med Res. 2008;128(4):335-52.
89. Soderland P, Lovekar S, Weiner DE, Brooks DR, Kaufman JS. Chronic kidney disease associated with environmental toxins and exposures. Adv Chronic Kidney Dis. 2010;17(3):254-64.

90. Jayasumana C, Gunatilake S, Senanayake P. Glyphosate, hard water and nephrotoxic metals: are they the culprits behind the epidemic of chronic kidney disease of unknown etiology in Sri Lanka? International journal of environmental research and public health. 2014;11(2):2125-47.
91. Siddarth M, Datta SK, Mustafa MD, Ahmed RS, Banerjee BD, Kalra OP, et al. Increased level of organochlorine pesticides in chronic kidney disease patients of unknown etiology: Role of GSTM1/GSTT1 polymorphism. Chemosphere. 2014;96:174-9.
92. Mohamed F, Endre ZH, Pickering JW, Jayamanne S, Palangasinghe C, Shahmy S, et al. Mechanism-specific injury biomarkers predict nephrotoxicity early following glyphosate surfactant herbicide (GPSH) poisoning. Toxicol Lett. 2016;258:1-10.
93. Wunnapuk K, Gobe G, Endre Z, Peake P, Grice JE, Roberts MS, et al. Use of a glyphosate-based herbicide-induced nephrotoxicity model to investigate a panel of kidney injury biomarkers. Toxicol Lett. 2014;225(1):192-200.
94. Tayeb W, Nakbi A, Trabelsi M, Miled A, Hammami M. Biochemical and histological evaluation of kidney damage after sub-acute exposure to 2,4-dichlorophenoxyacetic herbicide in rats: involvement of oxidative stress. Toxicology mechanisms and methods. 2012;22(9):696-704.
95. Valcke M, Levasseur ME, Soares da Silva A, Wesseling C. Pesticide exposures and chronic kidney disease of unknown etiology: an epidemiologic review. Environmental health : a global access science source. 2017;16(1):49.
96. Ritter L, Arbuckle TE. Can exposure characterization explain concurrence or discordance between toxicology and epidemiology? Toxicological sciences : an official journal of the Society of Toxicology. 2007;97(2):241-52.
97. Wanigasuriya KP, Peiris-John RJ, Wickremasinghe R. Chronic kidney disease of unknown aetiology in Sri Lanka: is cadmium a likely cause? BMC nephrology. 2011;12:32.
98. Stojsavljević A, Borković-Mitić S, Vujotić L, Grujičić D, Gavrović-Jankulović M, Manojlović D. The human biomonitoring study in Serbia: Background levels for arsenic, cadmium, lead, thorium and uranium in the whole blood of adult Serbian population. Ecotoxicology and Environmental Safety. 2019;169:402-9.
99. Balasooriya S, Munasinghe H, Herath AT, Diyabalanage S, Ileperuma OA, Manthrithilake H, et al. Possible links between groundwater geochemistry and chronic kidney disease of unknown etiology (CKDu): an investigation from the Ginnoruwa region in Sri Lanka. Exposure and Health. 2019.
100. Orr S.E BCC. Chronic Kidney Disease and Exposure to Nephrotoxic Metals. International journal of molecular sciences. 2017;18(1039).
101. Cohen SM, Arnold LL, Eldan M, Lewis AS, Beck BD. Methylated arsenicals: the implications of metabolism and carcinogenicity studies in rodents to human risk assessment. Critical reviews in toxicology. 2006;36(2):99-133.
102. Wilkinson SP, McHugh P, Horsley S, Tubbs H, Lewis M, Thould A, et al. Arsine toxicity aboard the Asiafreighter. British Medical Journal. 1975;3(5983):559-63.
103. Zheng LY, Umans JG, Tellez-Plaza M, Yeh F, Francesconi KA, Goessler W, et al. Urine arsenic and prevalent albuminuria: evidence from a population-based study. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2013;61(3):385-94.
104. Hsu LI, Hsieh FI, Wang YH, Lai TS, Wu MM, Chen CJ, et al. Arsenic Exposure From Drinking Water and the Incidence of CKD in Low to Moderate Exposed Areas of Taiwan: A 14-

Year Prospective Study. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2017;70(6):787-97.
105. Robles-Osorio ML, Sabath-Silva E, Sabath E. Arsenic-mediated nephrotoxicity. Renal failure. 2015;37(4):542-7.
106. Hsueh YM, Chung CJ, Shiue HS, Chen JB, Chiang SS, Yang MH, et al. Urinary arsenic species and CKD in a Taiwanese population: a case-control study. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2009;54(5):859-70.
107. Chen JW, Chen HY, Li WF, Liou SH, Chen CJ, Wu JH, et al. The association between total urinary arsenic concentration and renal dysfunction in a community-based population from central Taiwan. Chemosphere. 2011;84(1):17-24.
108. Chiou JM, Wang SL, Chen CJ, Deng CR, Lin W, Tai TY. Arsenic ingestion and increased microvascular disease risk: observations from the south-western arseniasis-endemic area in Taiwan. Int J Epidemiol. 2005;34(4):936-43.
109. Liu J, Liu Y, Habeebu SM, Waalkes MP, Klaassen CD. Chronic combined exposure to cadmium and arsenic exacerbates nephrotoxicity, particularly in metallothionein-I/II null mice. Toxicology. 2000;147(3):157-66.
110. Li Z, Piao F, Liu S, Wang Y, Qu S. Subchronic exposure to arsenic trioxide-induced oxidative DNA damage in kidney tissue of mice. Experimental and toxicologic pathology : official journal of the Gesellschaft fur Toxikologische Pathologie. 2010;62(5):543-7.
111. Johri N, Jacquillet G, Unwin R. Heavy metal poisoning: the effects of cadmium on the kidney. Biometals. 2010;23(5):783-92.
112. Thévenod F. Nephrotoxicity and the Proximal Tubule. Nephron Physiology. 2003;93(4):p87-p93.
113. Nordberg GF, Goyer R, Nordberg M. Comparative toxicity of cadmium-metallothionein and cadmium chloride on mouse kidney. Archives of pathology. 1975;99(4):192-7.
114. Girijashanker K, He L, Soleimani M, Reed JM, Li H, Liu Z, et al. Slc39a14 gene encodes ZIP14, a metal/bicarbonate symporter: similarities to the ZIP8 transporter. Molecular pharmacology. 2008;73(5):1413-23.
115. Järup L, Persson B, Elinder CG. Decreased glomerular filtration rate in solderers exposed to cadmium. Occupational and environmental medicine. 1995;52(12):818.
116. Brzoska MM, Kaminski M, Dziki M, Moniuszko-Jakoniuk J. Changes in the structure and function of the kidney of rats chronically exposed to cadmium. II. Histoenzymatic studies. Arch Toxicol. 2004;78(4):226-31.
117. Galazyn-Sidorczuk M, Brzoska MM, Jurczuk M, Moniuszko-Jakoniuk J. Oxidative damage to proteins and DNA in rats exposed to cadmium and/or ethanol. Chem Biol Interact. 2009;180(1):31-8.
118. Ferraro PM, Costanzi S, Naticchia A, Sturniolo A, Gambaro G. Low level exposure to cadmium increases the risk of chronic kidney disease: analysis of the NHANES 1999-2006. BMC Public Health. 2010;10(1):304.
119. Jarup L, Hellstrom L, Alfven T, Carlsson MD, Grubb A, Persson B, et al. Low level exposure to cadmium and early kidney damage: the OSCAR study. Occupational and environmental medicine. 2000;57(10):668-72.
120. Fowler BA, DuVal G. Effects of lead on the kidney: roles of high-affinity lead-binding proteins. Environmental Health Perspectives. 1991;91:77-80.
121. Ekong EB, Jaar BG, Weaver VM. Lead-related nephrotoxicity: a review of the epidemiologic evidence. Kidney international. 2006;70(12):2074-84.

122. Harari F, Sallsten G, Christensson A, Petkovic M, Hedblad B, Forsgard N, et al. Blood Lead Levels and Decreased Kidney Function in a Population-Based Cohort. American Journal of Kidney Diseases. 2018;72(3):381-9.
123. Swenberg JA, Short B, Borghoff S, Strasser J, Charbonneau M. The comparative pathobiology of α2u-globulin nephropathy. Toxicology and Applied Pharmacology. 1989;97(1):35-46.
124. Goyer RA. Mechanisms of lead and cadmium nephrotoxicity. Toxicology Letters. 1989;46(1):153-62.
125. Goyer RA. TOXIC AND ESSENTIAL METAL INTERACTIONS. Annual Review of Nutrition. 1997;17(1):37-50.
126. United States Geological Survey U. Water Hardness Classification of the United States Geological Survey United States: USGS; 2019 [Available from: https://water.usgs.gov/owq/hardness-alkalinity.html.
127. WHO WHO. Guidelines for Drinking-water Quality: Fourth edition incorporating the first addendum. Geneva, Switzerland: WHO; 2017.
128. Jayasumana C, Gajanayake R, Siribaddana S. Importance of Arsenic and pesticides in epidemic chronic kidney disease in Sri Lanka. BMC nephrology. 2014;15:124.
129. Wasana HMS, Aluthpatabendi D, Kularatne WMTD, Wijekoon P, Weerasooriya R, Bandara J. Drinking water quality and chronic kidney disease of unknown etiology (CKDu): synergic effects of fluoride, cadmium and hardness of water. Environmental geochemistry and health. 2016;38(1):157-68.
130. Rango T, Jeuland M, Manthrithilake H, McCornick P. Nephrotoxic contaminants in drinking water and urine, and chronic kidney disease in rural Sri Lanka. Science of the Total Environment. 2015;518-519:574-85.
131. Nishiya D, Omura T, Shimada K, Matsumoto R, Kusuyama T, Enomoto S, et al. Effects of erythropoietin on cardiac remodeling after myocardial infarction. Journal of pharmacological sciences. 2006;101(1):31-9.
132. Sanz AB, Santamaría B, Ruiz-Ortega M, Egido J, Ortiz A. Mechanisms of Renal Apoptosis in Health and Disease. Journal of the American Society of Nephrology. 2008;19(9):1634.
133. Nanayakkara S, Senevirathna ST, Parahitiyawa NB, Abeysekera T, Chandrajith R, Ratnatunga N, et al. Whole-exome sequencing reveals genetic variants associated with chronic kidney disease characterized by tubulointerstitial damages in North Central Region, Sri Lanka. Environmental health and preventive medicine. 2015;20(5):354-9.
134. Sayanthooran S, Magana-Arachchi DN, Gunerathne L, Abeysekera T. Potential diagnostic biomarkers for chronic kidney disease of unknown etiology (CKDu) in Sri Lanka: a pilot study. BMC nephrology. 2017;18(1):1-12.
135. Nanayakkara S, Senevirathna ST, Karunaratne U, Chandrajith R, Harada KH, Hitomi T, et al. Evidence of tubular damage in the very early stage of chronic kidney disease of uncertain etiology in the North Central Province of Sri Lanka: a cross-sectional study. Environmental health and preventive medicine. 2012;17(2):109-17.
136. Robinson WS. Ecological Correlations and the Behavior of Individuals. American Sociological Review. 1950;15(3):351-7.
137. Greenland S, Robins J. Invited Commentary: Ecologic Studies—Biases, Misconceptions, and Counterexamples. American journal of epidemiology. 1994;139(8):747-60.

138. Pearce N. The ecological fallacy strikes back. Journal of Epidemiology and Community Health. 2000;54(5):326.
139. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. American journal of public health. 1998;88(2):216-22.
140. González-Quiroz M, Camacho A, Faber D, Aragón A, Wesseling C, Glaser J, et al. Rationale, description and baseline findings of a community-based prospective cohort study of kidney function amongst the young rural population of Northwest Nicaragua. BMC nephrology. 2017;18(1):1-8.
141. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012;380(9859):2095-128.
142. Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1211-59.
143. Pearce N, Caplin B, Gunawardena N, Kaur P, O'Callaghan-Gordo C, Ruwanpathirana T. CKD of Unknown Cause: A Global Epidemic? Kidney International Reports. 2019;4(3):367-9.
144. Doll R, Muir CS, Waterhouse JAH. Incidence Tables. In: Doll R, Muir CS, Waterhouse JAH, editors. Cancer Incidence in Five Continents: Volume II – 1970. Berlin, Heidelberg: Springer Berlin Heidelberg; 1970. p. 85-332.
145. Doll R, Muir CS, Waterhouse JAH. Frequency of some Morphological Types of Neoplasm of Five Sites. In: Doll R, Muir CS, Waterhouse JAH, editors. Cancer Incidence in Five Continents: Volume II – 1970. Berlin, Heidelberg: Springer Berlin Heidelberg; 1970. p. 23-83.
146. McCredie M, editor Cancer Epidemiology in Migrant Populations. Genes and Environment in Cancer; 1998 1998//; Berlin, Heidelberg: Springer Berlin Heidelberg.
147. Nugent RA, Fathima SF, Feigl AB, Chyung D. The Burden of Chronic Kidney Disease on Developing Nations: A 21st Century Challenge in Global Health. Nephron Clinical Practice. 2011;118(3):c269-c77.
148. Thomas B, Wulf S, Bikbov B, Perico N, Cortinovis M, Courville de Vaccaro K, et al. Maintenance Dialysis throughout the World in Years 1990 and 2010. Journal of the American Society of Nephrology : JASN. 2015;26(11):2621-33.
149. World Health Organization W. Applying ICD-10 to verbal autopsy. Geneva, Switzerland: World Health Organization; 2016.
150. Organization WH. Global Health estimates 2016: deaths by cause, age, sex, by country and by region, 2000–2016. Geneva, Switzerland: World Health Organization; 2018.
151. Grams ME, Plantinga LC, Hedgeman E, Saran R, Myers GL, Williams DE, et al. Validation of CKD and related conditions in existing data sets: A systematic review. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2011;57(1):44-54.
152. Perneger TV, Klag MJ, Whelton PK. Cause of death in patients with end-stage renal disease: death certificates vs registry reports. American journal of public health. 1993;83(12):1735-8.
153. Rocco MV, Yan G, Gassman J, Lewis JB, Ornt D, Weiss B, et al. Comparison of causes of death using HEMO Study and HCFA end-stage renal disease death notification classification systems. The National Institutes of Health-funded Hemodialysis. Health Care Financing Administration. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2002;39(1):146-53.

154. Bello AK, Levin A, Tonelli M, Okpechi IG, Feehally J, Harris D, et al. Assessment of Global Kidney Health Care Status. Jama. 2017;317(18):1864-81.
155. Correa-Rotter R. The cost barrier to renal replacement therapy and peritoneal dialysis in the developing world. Peritoneal dialysis international : journal of the International Society for Peritoneal Dialysis. 2001;21.
156. Sinha A, Bagga A. Maintenance dialysis in developing countries. Pediatric Nephrology. 2015;30(2):211-9.
157. Ichiho HM, Anson R, Keller E, Lippwe K, Aitaoto N. An assessment of non-communicable diseases, diabetes, and related risk factors in the Federated States of Micronesia, State of Pohnpei: a systems perspective. Hawaii J Med Public Health. 2013;72(5 Suppl 1):49-56.
158. Anand S, Bitton A, Gaziano T. The Gap between Estimated Incidence of End-Stage Renal Disease and Use of Therapy. PloS one. 2013;8(8):e72860.
159. White SL, Chadban SJ, Jan S, Chapman JR, Cass A. How can we achieve global equity in provision of renal replacement therapy? Bull World Health Organ. 2008;86(3):229-37.
160. Global Burden of Disease Results Tool [Internet]. IHME. 2018 [cited 12/12/2019]. Available from: http://ghdx.healthdata.org/gbd-results-tool.
161. IBIS. Stanardized Mortality Ratio United Kingdom2016 [Available from: https://ibis.health.state.nm.us/resource/SMR_ISR.html#:~:text=using%20the%20formula%3A%20(C%20x,number%20of%20deaths%20was%2026.5.
162. Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio (SMR). American journal of epidemiology. 1990;131(2):373-5.
163. Lunyera J, Mohottige D, Von Isenburg M, Jeuland M, Patel UD, Stanifer JW. CKD of Uncertain Etiology: A Systematic Review. Clinical Journal of The American Society of Nephrology: CJASN. 2016;11(3):379-85.
164. Jayasekara KB, Dissanayake DM, Sivakanesan R, Ranasinghe A, Karunarathna RH, Priyantha Kumara GW. Epidemiology of chronic kidney disease, with special emphasis on chronic kidney disease of uncertain etiology, in the north central region of Sri Lanka. Journal of epidemiology. 2015;25(4):275-80.
165. Kerr PG, Tran HTB, Ha Phan H-A, Liew A, Hooi LS, Johnson DW, et al. Nephrology in the Oceania and South East Asia region: perspectives and challenges. Kidney international. 2018;94(3):465-70.
166. Weil EJ, Nelson RG. Kidney disease among the indigenous peoples of Oceania. Ethnicity & disease. 2006;16(2 Suppl 2):S2-24-30.
167. Garcia-Garcia G, Briseño-Rentería G, Luquín-Arellan VH, Gao Z, Gill J, Tonelli M. Survival among Patients with Kidney Failure in Jalisco, Mexico. Journal of the American Society of Nephrology. 2007;18(6):1922.
168. Stewart JH MM, McDonald SP. The incidence of treated end-stage renal disease in New Zealand Maori and Pacific Island People and in indigenous Australians. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2004;19(3):678–85.
169. Cobo G, Hecking M, Port FK, Exner I, Lindholm B, Stenvinkel P, et al. Sex and gender differences in chronic kidney disease: progression to end-stage renal disease and haemodialysis. Clinical science (London, England : 1979). 2016;130(14):1147-63.
170. Carrero JJ, Hecking M, Chesnaye NC, Jager KJ. Sex and gender disparities in the epidemiology and outcomes of chronic kidney disease. Nature Reviews Nephrology. 2018;14(3):151-64.

171. Glassock R, Delanaye P, El Nahas M. An Age-Calibrated Classification of Chronic Kidney Disease. Jama. 2015;314(6):559-60.
172. O'Hare AM, Choi AI, Bertenthal D, Bacchetti P, Garg AX, Kaufman JS, et al. Age affects outcomes in chronic kidney disease. Journal of the American Society of Nephrology : JASN. 2007;18(10):2758-65.
173. Inker LA, Shafi T, Okparavero A, Tighiouart H, Eckfeldt JH, Katz R, et al. Effects of Race and Sex on Measured GFR: The Multi-Ethnic Study of Atherosclerosis. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2016;68(5):743-51.
174. Eastwood JB, Kerry SM, Plange-Rhule J, Micah FB, Antwi S, Boa FG, et al. Assessment of GFR by four methods in adults in Ashanti, Ghana: the need for an eGFR equation for lean African populations. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2010;25(7):2178-87.
175. Herrera-Valdes R, Almaguer-Lopez M. Strategies for national health care systems and centers in the emerging world: Central America and the Caribbean—The case of Cuba. Kidney international. 2005;68:S66-S8.
176. Soyibo AK, Barton EN. Report from the Caribbean renal registry, 2006. The West Indian medical journal. 2007;56(4):355-63.
177. Garcia Garcia G, Gutierrez Padilla JA, Perez-Gomez H, Chávez-Iñiguez J, Morraz-Mejia E, Amador-Jimenez M, et al. Chronic kidney disease of unknown cause in Mexico: The case of Poncitlan, Jalisco. Clinical Nephrology. 2019.
178. Cusumano AM, Rosa-Diez GJ, Gonzalez-Bedat MC. Latin American Dialysis and Transplant Registry: Experience and contributions to end-stage renal disease epidemiology. World J Nephrol. 2016;5(5):389-97.
179. Orantes CM, Herrera R, Almaguer M, Brizuela EG, Nunez L, Alvarado NP, et al. Epidemiology of chronic kidney disease in adults of Salvadoran agricultural communities. MEDICC review. 2014;16(2):23-30.
180. Ceron A, Fort MP, Morine CM, Lou-Meda R. Chronic kidney disease among children in Guatemala. Revista panamericana de salud publica = Pan American journal of public health. 2014;36(6):376-82.
181. Butler-Dawson J, Krisher L, Asensio C, Cruz A, Tenney L, Weitzenkamp D, et al. Risk Factors for Declines in Kidney Function in Sugarcane Workers in Guatemala. Journal of Occupational and Environmental Medicine. 2018;60(6).
182. Valdés RH, Orantes CM, López MA, Marín LL, Arévalo PA, Smith González MJ, et al. Clinical characteristics of chronic kidney disease of non-traditional causes in women of agricultural communities in El Salvador. Clinical Nephrology. 2015;83:56-63.
183. Cusumano A, Garcia-Garcia G, Di Gioia C, Hermida O, Lavorato C, Agost Carreño C, et al. End-Stage Renal Disease and Its Treatment in Latin America in the Twenty-First Century. Renal failure. 2006;28(8):631-7.
184. George C, Mogueo A, Okpechi I, Echouffo-Tcheugui JB, Kengne AP. Chronic kidney disease in low-income to middle-income countries: the case for increased screening. BMJ Global Health. 2017;2(2):e000256.
185. International Diabetes Federation Diabetes Atlas [Internet]. 2020 [cited 18/03/2020]. Available from: https://diabetesatlas.org/data/en/compare/143-127/idf-country-data-comparision.html.

186. Wijkstrom J, Leiva R, Elinder CG, Leiva S, Trujillo Z, Trujillo L, et al. Clinical and pathological characterization of mesoamerican nephropathy: A new kidney disease in central america. American Journal of Kidney Diseases. 2013;62(5):908-18.
187. Pecoits-Filho R, Sola L, Correa-Rotter R, Claure-Del Granado R, Douthat WG, Bellorin-Font E. Kidney disease in Latin America: current status, challenges, and the role of the ISN in the development of nephrology in the region. Kidney international. 2018;94(6):1069-72.
188. Ordunez P, Martinez R, Reveiz L, Chapman E, Saenz C, Soares da Silva A, et al. Chronic kidney disease epidemic in Central America: urgent public health action is needed amid causal uncertainty. PLoS neglected tropical diseases. 2014;8(8):e3019.
189. Barsoum RS. Burden of chronic kidney disease: North Africa. Kidney International Supplements. 2013;3(2):164-6.
190. Seck SM, Doupa D, Gueye L, Dia CA. Prevalence of chronic kidney disease and associated factors in senegalese populations: a community-based study in saint-louis. Nephro-urology monthly. 2014;6(5):e19085-e.
191. Stanifer JW, Jing B, Tolan S, Helmke N, Mukerjee R, Naicker S, et al. The epidemiology of chronic kidney disease in sub-Saharan Africa: a systematic review and meta-analysis. The Lancet Global Health. 2014;2(3):e174-e81.
192. Hutchinson A. HIV/AIDS in Europe: Moving from Death Sentence to Chronic Disease Management. Conscience. 2006;27(2):48.
193. Dwyer-Lindgren L, Cork MA, Sligar A, Steuben KM, Wilson KF, Provost NR, et al. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. Nature. 2019;570(7760):189-93.
194. UNAIDS Lesotho [Internet]. United Nations. 2020 [cited 19/03/2020]. Available from: https://www.unaids.org/en/regionscountries/countries/lesotho.
195. UNAIDS Eswatini [Internet]. United Nations. 2020 [cited 19/03/2020]. Available from: https://www.unaids.org/en/regionscountries/countries/swaziland.
196. HIV and AIDS in Lesotho [Internet]. Avert. 2018. Available from: https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/lesotho.
197. HIV and AIDS in Eswatini [Internet]. Avert. 2020 [cited 19/03/2020]. Available from: https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/swaziland.
198. Abd ElHafeez S, Bolignano D, D'Arrigo G, Dounousi E, Tripepi G, Zoccali C. Prevalence and burden of chronic kidney disease among the general population and high-risk groups in Africa: a systematic review. BMJ open. 2018;8(1):e015069.
199. Imtiaz S, Salman B, Qureshi R, Drohlia M, Ahmad A. A review of the epidemiology of chronic kidney disease in Pakistan: A global and regional perspective. Saudi Journal of Kidney Diseases and Transplantation. 2018;29(6):1441-51.
200. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Research and Clinical Practice. 2014;103(2):137-49.
201. Ramachandran A, Snehalatha C, Ma RCW. Diabetes in South-East Asia: An update. Diabetes Research and Clinical Practice. 2014;103(2):231-7.
202. Jha V. Current status of end-stage renal disease care in India and Pakistan. Kidney International Supplements. 2013;3(2):157-60.
203. Akhtar S, Nasir JA, Abbas T, Sarwar A. Diabetes in Pakistan: A systematic review and meta-analysis. Pak J Med Sci. 2019;35(4):1173-8.
204. Rizvi S, Manzoor K. Causes of Chronic Renal Failure in Pakistan: A Single Large Center Experience. Saudi Journal of Kidney Diseases and Transplantation. 2002;13(3):376-9.

205. Pringle D. Mapping Disease Risk Estimates Based on Small Numbers: An Assessment of Empirical Bayes Techniques*. Economic and Social Review. 1996.
206. Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. Environmental health perspectives. 2004;112(9):1016-25.
207. Wakefield J. Disease mapping and spatial regression with count data. Biostatistics. 2007;8(2):158-83.
208. Heinzl H, Waldhoer T. Relevance of the type III error in epidemiological maps. Int J Health Geogr. 2012;11:34-.
209. Alhdiri MAS, Samat NA, Mohamed Z. Disease Mapping for Stomach Cancer in Libya Based on Besag– York– Mollié (BYM) Model. Asian Pac J Cancer Prev. 2017;18(6):1479-84.
210. Horsky J, Drucker EA, Ramelson HZ. Accuracy and Completeness of Clinical Coding Using ICD-10 for Ambulatory Visits. AMIA Annu Symp Proc. 2018;2017:912-20.
211. Abraham G, Varughese S, Thandavan T, Iyengar A, Fernando E, Naqvi SA, et al. Chronic kidney disease hotspots in developing countries in South Asia. Clinical kidney journal. 2016;9(1):135-41.
212. Agriculture ICoFa. Report on Haryana agriculture and farmers' welfare. New Delhi: IFCA; 2017.
213. Wanigasuriya KP, Peiris-John RJ, Wickremasinghe R, Hittarage A. Chronic renal failure in North Central Province of Sri Lanka: an environmentally induced disease. Trans R Soc Trop Med Hyg. 2007;101.
214. Weaver VM, Fadrowski JJ, Jaar BG. Global dimensions of chronic kidney disease of unknown etiology (CKDu): A modern era environmental and/or occupational nephropathy? BMC nephrology. 2015;16 (1) (no pagination)(145).
215. Chandrajith R, Dissanayake CB, Ariyarathna T, Herath HM, Padmasiri JP. Dose-dependent Na and Ca in fluoride-rich drinking water--another major cause of chronic renal failure in tropical arid regions. The Science of the total environment. 2011;409(4):671-5.
216. UNFPA UNPFSL. Census of Population and Housing 2012 Sri Lanka. 2012.
217. Bank TW. Urban population growth (annual %) - Sri Lanka USA2019 [Available from: https://data.worldbank.org/indicator/SP.URB.GROW?locations=LK.
218. Spatial Data Download: Sri Lanka [Internet]. DIVA-GIS. 2011. Available from: https://www.diva-gis.org/datadown.
219. Ministry of Home Affairs GoI. Census 2011. 2011.
220. Mohan S, Jarhyan P, Ghosh S, Venkateshmurthy NS, Gupta R, Rana R, et al. UDAY: A comprehensive diabetes and hypertension prevention and management program in India. BMJ open. 2018;8(6):e015919-e.
221. India shapefile [Internet]. 2014. Available from: https://www.arcgis.com/home/item.html?id=cf9b387de48248a687aafdd4cdff1127.
222. Nair M, Ali MK, Ajay VS, Shivashankar R, Mohan V, Pradeepa R, et al. CARRS Surveillance study: design and methods to assess burdens from multiple perspectives. BMC Public Health. 2012;12(1):701.
223. Chabot-Couture G, Nigmatulina K, Eckhoff P. An Environmental Data Set for Vector-Borne Disease Modeling and Epidemiology. PloS one. 2014;9(4):e94741.
224. Wrable M, Kulinkina AV, Liss A, Koch M, Cruz MS, Biritwum N-K, et al. The use of remotely sensed environmental parameters for spatial and temporal schistosomiasis prediction across climate zones in Ghana. Environmental Monitoring and Assessment. 2019;191(2):301.

225. Hu Y, Xia C, Li S, Ward MP, Luo C, Gao F, et al. Assessing environmental factors associated with regional schistosomiasis prevalence in Anhui Province, Peoples' Republic of China using a geographical detector method. Infectious Diseases of Poverty. 2017;6(1):87.
226. Minale AS, Alemu K. Mapping malaria risk using geographic information systems and remote sensing: The case of Bahir Dar City, Ethiopia. Geospatial Health. 2018;13(1).
227. Forecasts ECfM-RW. ERA5 Reanalysis Dataset Unkted Kingdom: ECMWF; 2019 [Available from: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.
228. USGS EROS Archive - Digital Elevation - Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global [Internet]. United States Geological Survey, (USGS). 2019 [cited 20/11/2019]. Available from: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1-arc?qt-science_center_objects=0#qt-science_center_objects.
229. European Space Agency Climate Change Initiative E-C. ESA-CCI Land cover maps The Netherlands: ESA-CCI; 2015 [Available from: https://www.esa-landcover-cci.org/?q=node/158.
230. Hass AL, Ellis KN, Reyes Mason L, Hathaway JM, Howe DA. Heat and Humidity in the City: Neighborhood Heat Index Variability in a Mid-Sized City in the Southeastern United States. International journal of environmental research and public health. 2016;13(1).
231. Quinn A, Kinney P, Shaman J. Predictors of summertime heat index levels in New York City apartments. Indoor Air. 2017;27(4):840-51.
232. Steadman RG. The Assessment of Sultriness. Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science. Journal of Applied Meteorology. 1979;18(7):861-73.
233. Steadman RG. The Assessment of Sultriness. Part II: Effects of Wind, Extra Radiation and Barometric Pressure on Apparent Temperature. Journal of Applied Meteorology. 1979;18(7):874-85.
234. Fletcher BA, Lin S, Fitzgerald EF, Hwang S-A. Association of Summer Temperatures With Hospital Admissions for Renal Diseases in New York State: A Case-Crossover Study. American journal of epidemiology. 2012;175(9):907-16.
235. Burkart K, Schneider A, Breitner S, Khan MH, Krämer A, Endlicher W. The effect of atmospheric thermal conditions and urban thermal pollution on all-cause and cardiovascular mortality in Bangladesh. Environmental Pollution. 2011;159(8):2035-43.
236. Barnett AG, Tong S, Clements ACA. What measure of temperature is the best predictor of mortality? Environmental research. 2010;110(6):604-11.
237. Meteorology AMCo. Integrated African strategy on Meteorology (Weather and Climate Services). World Meteorological Organization; 2013.
238. Anderson GB, Bell ML, Peng RD. Methods to calculate the heat index as an exposure metric in environmental health research. Environmental health perspectives. 2013;121(10):1111-9.
239. Shtiliyanova A, Bellocchi G, Borras D, Eza U, Martin R, Carrère P. Kriging-based approach to predict missing air temperature data. Computers and Electronics in Agriculture. 2017;142:440-9.
240. Mazzella A, Mazzella A. The Importance of the Model Choice for Experimental Semivariogram Modeling and Its Consequence in Evaluation Process. Journal of Engineering. 2013;2013:960105.

241. Addis H.K KA, Strohmeier S. Spatial Variability of Selected Soil Attributes under Agricultural Land Use System in a Mountainous Watershed, Ethiopia. International Journal of Geosciences. 2015;6:605-13.
242. ESRI. Fitting a model to the empirical semivariogram United States: ESRI; 2019 [Available from: https://desktop.arcgis.com/es/arcmap/latest/extensions/geostatistical-analyst/fitting-a-model-to-the-empirical-semivariogram.htm.
243. Krivoruchko K, Gribov A. Evaluation of empirical Bayesian kriging. Spatial Statistics. 2019;32:100368.
244. Gribov A, Krivoruchko K. Empirical Bayesian kriging implementation and usage. Science of The Total Environment. 2020;722:137290.
245. Institute S. Log transformations: how to handle negative values USA2018 [Available from: https://blogs.sas.com/content/iml/2011/04/27/log-transformations-how-to-handle-negative-data-values.html.
246. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography. 2017;40(8):913-29.
247. Maleika W. Inverse distance weighting method optimization in the process of digital terrain model creation based on data collected from a multibeam echosounder. Applied Geomatics. 2020;12(4):397-407.
248. ESRI. Resample (Data Management) California, USA2020 [Available from: https://pro.arcgis.com/en/pro-app/tool-reference/data-management/resample.htm.
249. Elkhrachy I. Vertical accuracy assessment for SRTM and ASTER Digital Elevation Models: A case study of Najran city, Saudi Arabia. Ain Shams Engineering Journal. 2018;9(4):1807-17.
250. Wesseling C, Hogstedt C, Picado A, Johansson L. Unintentional fatal paraquat poisonings among agricultural workers in Costa Rica: report of 15 cases. American journal of industrial medicine. 1997;32(5):433-41.
251. Classification Concepts: Land Cover Classification System [Internet]. FAO. 2016.
252. Institute E. Specify a coordinate system California, USA2019 [Available from: https://pro.arcgis.com/en/pro-app/help/mapping/properties/specify-a-coordinate-system.htm.
253. Hedeker D, Gibbons RD, Flay BR. Random-effects regression models for clustered data with an example from smoking prevention research. Journal of consulting and clinical psychology. 1994;62(4):757-65.
254. Cao B, Quan X, Brown N, Stewart-Jones E, Gruber S. GlobSim (v1.0): deriving meteorological time series for point locations from multiple global reanalyses. Geosci Model Dev. 2019;12(11):4661-79.
255. Dodson R MD. Daily air temperature interpolated at high spatial resolution over a large mountainous region Climate Research. 1997;8:1-20.
256. O'Donnell JK, Tobey M, Weiner DE, Stevens LA, Johnson S, Stringham P, et al. Prevalence of and risk factors for chronic kidney disease in rural Nicaragua. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2011;26(9):2798-805.
257. Laux TS, Bert PJ, Barreto Ruiz GM, González M, Unruh M, Aragon A, et al. Nicaragua revisited: evidence of lower prevalence of chronic kidney disease in a high-altitude, coffee-growing village. Journal of nephrology. 2012;25(4):533-40.

258. Kaur P, Gunawardena N, Kumaresan J. A review of chronic kidney disease of unknown etiology in Sri Lanka, 2001–2015. Indian journal of nephrology. 2019.
259. Klinkenberg E, van der Hoek W, Amerasinghe FP. A malaria risk analysis in an irrigated area in Sri Lanka. Acta Tropica. 2004;89(2):215-25.
260. Athapattu BCL, Thalgaspitiya TWLR, Yasaratne ULS, Vithanage M. Biochar-based constructed wetlands to treat reverse osmosis rejected concentrates in chronic kidney disease endemic areas in Sri Lanka. Environmental geochemistry and health. 2017;39(6):1397-407.
261. Jayasumana C, Ranasinghe O, Ranasinghe S, Siriwardhana I, Gunatilake S, Siribaddana S. Reverse osmosis plant maintenance and efficacy in chronic kidney disease endemic region in Sri Lanka. Environmental health and preventive medicine. 2016;21(6):591-6.
262. Price AJ, Crampin AC, Amberbir A, Kayuni-Chihana N, Musicha C, Tafatatha T, et al. Prevalence of obesity, hypertension, and diabetes, and cascade of care in sub-Saharan Africa: a cross-sectional, population-based study in rural and urban Malawi. The Lancet Diabetes & Endocrinology. 2018;6(3):208-22.
263. Crampin AC, Kayuni N, Amberbir A, Musicha C, Koole O, Tafatatha T, et al. Hypertension and diabetes in Africa: design and implementation of a large population-based study of burden and risk factors in rural and urban Malawi. Emerging themes in epidemiology [Internet]. 2016 2016; 13:[3 p.]. Available from: http://europepmc.org/abstract/MED/26839575

http://europepmc.org/articles/PMC4736489?pdf=render

http://europepmc.org/articles/PMC4736489

https://doi.org/10.1186/s12982-015-0039-2.
264. Crampin AC, Dube A, Mboma S, Price A, Chihana M, Jahn A, et al. Profile: the Karonga Health and Demographic Surveillance System. Int J Epidemiol. 2012;41(3):676-85.
265. WHO WHO. Global Physical Activity Questionnaire (GPAQ) Geneva: World Health Organization; 2017.
266. Country factsheets: Malawi HIV and AIDS Estimates, 2018 [Internet]. 2019. Available from: https://www.unaids.org/en/regionscountries/countries/malawi.
267. Izzedine H, Harris M, Perazella MA. The nephrotoxic effects of HAART. Nature Reviews Nephrology. 2009;5(10):563-73.
268. Fabian J, George JA, Etheredge HR, van Deventer M, Kalyesubula R, Wade AN, et al. Methods and reporting of kidney function: a systematic review of studies from sub-Saharan Africa. Clinical kidney journal. 2019;12(6):778-87.
269. Coresh J, Astor BC, McQuillan G, Kusek J, Greene T, Van Lente F, et al. Calibration and random variation of the serum creatinine assay as critical elements of using equations to estimate glomerular filtration rate. American Journal of Kidney Diseases. 2002;39(5):920-9.
270. 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. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2015;30 Suppl 4(Suppl 4):iv6-16.
271. K/DOQI. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2002;39(2 Suppl 1):S1-266.

272. Bukabau JB, Sumaili EK, Cavalier E, Pottel H, Kifakiou B, Nkodila A, et al. Performance of glomerular filtration rate estimation equations in Congolese healthy adults: The inopportunity of the ethnic correction. PloS one. 2018;13(3):e0193384-e.
273. van Deventer HE, George JA, Paiker JE, Becker PJ, Katz IJ. Estimating glomerular filtration rate in black South Africans by use of the modification of diet in renal disease and Cockcroft-Gault equations. Clinical chemistry. 2008;54(7):1197-202.
274. Wyatt CM, Schwartz GJ, Owino Ong'or W, Abuya J, Abraham AG, Mboku C, et al. Estimating kidney function in HIV-infected adults in Kenya: comparison to a direct measure of glomerular filtration rate by iohexol clearance. PloS one. 2013;8(8):e69601.
275. Moodley N, Hariparshad S, Peer F, Gounden V. Evaluation of the CKD-EPI creatinine based glomerular filtration rate estimating equation in Black African and Indian adults in KwaZulu-Natal, South Africa. Clinical biochemistry. 2018;59:43-9.
276. Glaser N, Deckert A, Phiri S, Rothenbacher D, Neuhann F. Comparison of Various Equations for Estimating GFR in Malawi: How to Determine Renal Function in Resource Limited Settings? PloS one. 2015;10(6):e0130453.
277. Stanifer JW, Maro V, Egger J, Karia F, Thielman N, Turner EL, et al. The epidemiology of chronic kidney disease in Northern Tanzania: a population-based survey. PloS one. 2015;10(4):e0124506-e.
278. Association TR. CKD Stages 2019 [Available from: https://renal.org/information-resources/the-uk-eckd-guide/ckd-stages/.
279. Mula-Abed W-AS, Al Rasadi K, Al-Riyami D. Estimated Glomerular Filtration Rate (eGFR): A Serum Creatinine-Based Test for the Detection of Chronic Kidney Disease and its Impact on Clinical Practice. Oman Med J. 2012;27(2):108-13.
280. National Institute for Health and Care Excellence N. Chronic kidney disease in adults: assessment and management. National Institute for Health and Care Excellence; 2014.
281. Greenland S, Daniel R, Pearce N. Outcome modelling strategies in epidemiology: traditional methods and basic alternatives. International journal of epidemiology. 2016;45(2):565-75.
282. Herrington WG, Smith M, Bankhead C, Matsushita K, Stevens S, Holt T, et al. Body-mass index and risk of advanced chronic kidney disease: Prospective analyses from a primary care cohort of 1.4 million adults in England. PloS one. 2017;12(3):e0173515-e.
283. Bailey PK, Tomson CR, Kinra S, Ebrahim S, Radhakrishna KV. Differences in estimation of creatinine generation between renal function estimating equations in an Indian population: cross-sectional data from the Hyderabad arm of the Indian migration study. BMC nephrology. 2013;14.
284. Afsar B, Elsurer Afsar R, Kanbay A, Covic A, Ortiz A, Kanbay M. Air pollution and kidney disease: review of current evidence. Clinical kidney journal. 2019;12(1):19-32.
285. Al-Aly Z, Bowe B. Air Pollution and Kidney Disease. Clinical Journal of the American Society of Nephrology. 2020;15(3):301.
286. Jackson WA, Pry JM, Wickramasinghe S, Rupasinghe R, Jayawickrama S, Badurdeen Z, et al. Utilizing a one health approach for identifying risk factors associated with an epidemic of chronic kidney disease of unknown etiology (CKDu) in the North Central region of Sri Lanka. Annals of Global Health. 2016;82 (3):575-6.
287. Valcke M, Levasseur M-E, Soares da Silva A, Wesseling C. Pesticide exposures and chronic kidney disease of unknown etiology: an epidemiologic review. Environmental health : a global access science source. 2017;16(1):49-.

288. Pearce N, Caplin B. Let's take the heat out of the CKDu debate: more evidence is needed. Occupational and environmental medicine. 2019;76(6):357-9.
289. Roncal-Jimenez C, Garcia-Trabanino R, Barregard L, Lanaspa MA, Wesseling C, Harra T, et al. Heat Stress Nephropathy From Exercise-Induced Uric Acid Crystalluria: A Perspective on Mesoamerican Nephropathy. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2016;67(1):20-30.
290. Lewin PA, Fisher M, Weber B. Do rainfall conditions push or pull rural migrants: evidence from Malawi. Agricultural Economics. 2012;43(2):191-204.
291. Mtika MM. Political economy, labor migration, and the AIDS epidemic in rural Malawi. Social science & medicine (1982). 2007;64(12):2454-63.
292. Food and Agriculture Organization of the United Nations F. Family Farming Knowledge Platform. Washington, USA; 2011.
293. Julien JC, Bravo-Ureta BE, Rada NE. Assessing farm performance by size in Malawi, Tanzania, and Uganda. Food Policy. 2019;84:153-64.
294. Negatu B, Kromhout H, Mekonnen Y, Vermeulen R. Use of Chemical Pesticides in Ethiopia: A Cross-Sectional Comparative Study on Knowledge, Attitude and Practice of Farmers and Farm Workers in Three Farming Systems. The Annals of Occupational Hygiene. 2016;60(5):551-66.
295. Kasambala Donga T, Eklo OM. Environmental load of pesticides used in conventional sugarcane production in Malawi. Crop Protection. 2018;108:71-7.
296. Nanayakkara S, Senevirathna ST, Abeysekera T, Chandrajith R, Ratnatunga N, Gunarathne ED, et al. An integrative study of the genetic, social and environmental determinants of chronic kidney disease characterized by tubulointerstitial damages in the North Central Region of Sri Lanka. Journal of occupational health. 2014;56(1):28-38.
297. Dharma-Wardana MW, Amarasiri SL, Dharmawardene N, Panabokke CR. Chronic kidney disease of unknown aetiology and ground-water ionicity: study based on Sri Lanka. Environmental geochemistry and health. 2015;37(2):221-31.
298. Bowling A. Mode of questionnaire administration can have serious effects on data quality. Journal of Public Health. 2005;27(3):281-91.
299. Manski CF, Molinari F. SKIP SEQUENCING: A DECISION PROBLEM IN QUESTIONNAIRE DESIGN. Ann Appl Stat. 2008;2(1):264-85.
300. Liang G, Pan L, Liu X. Assessment of Typical Heavy Metals in Human Hair of Different Age Groups and Foodstuffs in Beijing, China. International journal of environmental research and public health. 2017;14(8):914.
301. Pereira R, Ribeiro R, Gonçalves F. Scalp hair analysis as a tool in assessing human exposure to heavy metals (S. Domingos mine, Portugal). The Science of the total environment. 2004;327(1-3):81-92.
302. Schultze B, Lind PM, Larsson A, Lind L. Whole blood and serum concentrations of metals in a Swedish population-based sample. Scandinavian Journal of Clinical and Laboratory Investigation. 2014;74(2):143-8.
303. Čurdová E, Koplík R, Matoušek JP, Suchánek M. Direct determination of cadmium in urine by electrothermal atomic absorption spectrometry after in situ electrodeposition. Talanta. 2005;67(5):926-32.
304. Lemos VA, de Carvalho AL. Determination of cadmium and lead in human biological samples by spectrometric techniques: a review. Environmental Monitoring and Assessment. 2010;171(1):255-65.

305. Registry AfTSaD. Case Studies in Environmental Medicine (CSEM) Cadmium Toxicity. United States; 2011.
306. Water NRCUSoAiD. Biomarkers of Arsenic Exposure. Arsenic in Drinking Water. Washington (DC): National Academies Press (US); 1999.
307. Omuto CT VR. Soil Nutrient Loss Assessment in Malawi: Technical Report. 2018.
308. Food and Agriculture Organization of the United Nations F. FAO Soils Portal United States: Food and Agriculture Organization of the United Nations, (FAO); 1974 [Available from: http://www.fao.org/soils-portal/soil-survey/soil-classification/fao-legend/key-to-the-fao-soil-units/en/.
309. Ferrari CR, do Nascimento HdAF, Rodgher S, Almeida T, Bruschi AL, Nascimento MRLd, et al. Effects of the discharge of uranium mining effluents on the water quality of the reservoir: an integrative chemical and ecotoxicological assessment. Scientific Reports. 2017;7(1):13919.
310. MoIaWD GotRoM. Annexure 3A: Borehole water monitoring census report: fielw review of 35 water monitoring boreholes: final report. Ministry of Irrigation and Water Development;; 2011.
311. Network RWS. Afridev Handpump Specification. Switzerland; 2007.
312. Dijkshoorn JA LJ, Huting JHM, Kempen B. Soil and terrain database of the Republic of Malawi. Report 2016/01. Wageningen; 2016.
313. Tiwale S RM, Zwarteveen M. The Power of Pipes: Mapping Urban Water Inequities through the Material Properties of Networked Water Infrastructures - The Case of Lilongwe, Malawi. Water Alternatives. 2018;11.
314. S2 prototype land cover 20m map of Africa 2016 [Internet]. European Space Agency Climate Change Initiative, ESA-CCI. 2017 [cited 20/04/2020]. Available from: http://2016africalandcover20m.esrin.esa.int/.
315. Standards MBo. Drinking water – Specification. 2013.
316. Jolliffe I. Principal component analysis. 2nd ed. http://lst-iiepiiep-unescoorg/cgi-bin/wwwi32exe/[in=epidoc1in]/?t2000=017716/(100). 2005;98.
317. Shah-Kulkarni S, Lee S, Jeong KS, Hong Y-C, Park H, Ha M, et al. Prenatal exposure to mixtures of heavy metals and neurodevelopment in infants at 6 months. Environmental research. 2020;182:109122.
318. Bobb JF, Valeri L, Claus Henn B, Christiani DC, Wright RO, Mazumdar M, et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics. 2014;16(3):493-508.
319. Valeri L, Mazumdar MM, Bobb JF, Claus Henn B, Rodrigues E, Sharif OIA, et al. The Joint Effect of Prenatal Exposure to Metal Mixtures on Neurodevelopmental Outcomes at 20-40 Months of Age: Evidence from Rural Bangladesh. Environmental health perspectives. 2017;125(6):067015-.
320. Sheth AN, Russo ET, Menon M, Wannemuehler K, Weinger M, Kudzala AC, et al. Impact of the integration of water treatment and handwashing incentives with antenatal services on hygiene practices of pregnant women in Malawi. The American journal of tropical medicine and hygiene. 2010;83(6):1315-21.
321. Lim JT, Tan YQ, Valeri L, Lee J, Geok PP, Chia SE, et al. Association between serum heavy metals and prostate cancer risk – A multiple metal analysis. Environment International. 2019;132:105109.
322. Popenda A. Effect of redox potential on heavy metals and As behavior in dredged sediments. Desalination and Water Treatment. 2014;52.

323. Tack FM, Callewaert OWJJ, Verloo MG. Metal solubility as a function of pH in a contaminated, dredged sediment affected by oxidation. Environmental Pollution. 1996;91(2):199-208.
324. Tchounwou PB, Yedjou CG, Patlolla AK, Sutton DJ. Heavy metal toxicity and the environment. Exp Suppl. 2012;101:133-64.
325. Braeken J vAM. An empirical Kaiser criterion. Psychol Methods. 2017;22(3):450-66.
326. Lopez-Marin L, Chavez Y, Garcia XA, Flores WM, Garcia YM, Herrera R, et al. Histopathology of chronic kidney disease of unknown etiology in Salvadoran agricultural communities. MEDICC review. 2014;16(2):49-54.
327. Bandara JM, Wijewardena HV, Liyanege J, Upul MA, Bandara JM. Chronic renal failure in Sri Lanka caused by elevated dietary cadmium: Trojan horse of the green revolution. Toxicol Lett. 2010;198(1):33-9.
328. Anonymous. Erratum: Heat stress, hydration and uric acid: A crosssectional study in workers of three occupations in a hotspot of Mesoamerican nephropathy in Nicaragua (BMJ Open (2016) 6 (e011034) DOI: 10.1136/bmjopen-2016-011034). BMJ open. 2017;7 (1) (no pagination)(e011034corr1).
329. Marill KA. Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression. Academic Emergency Medicine. 2004;11(1):94-102.
330. Prozialeck WC, Edwards JR. Mechanisms of cadmium-induced proximal tubule injury: new insights with implications for biomonitoring and therapeutic interventions. J Pharmacol Exp Ther. 2012;343(1):2-12.
331. Sanders A, Mazzella M, Malin A, Hair G, Busgang S, Saland J, et al. Combined exposure to lead, cadmium, mercury, and arsenic and kidney health in adolescents age 12–19 in NHANES 2009–2014. Environment International. 2019;131.
332. Price RG. Cadmium Nephropathy and Smoking. Clinical Medicine Insights: Urology. 2017;10:1179561117726090.
333. Bernard A, Lauwerys R. Early markers of cadmium nephrotoxicity: Biological significance and predictive value. Toxicological & Environmental Chemistry. 1990;27(1-3):65-72.
334. Wallin M, Sallsten G, Lundh T, Barregard L. Low-level cadmium exposure and effects on kidney function. Occupational and environmental medicine. 2014;71(12):848-54.
335. Kang M, Tian Y, Peng S, Wang M. Effect of dissolved oxygen and nutrient levels on heavy metal contents and fractions in river surface sediments. Science of The Total Environment. 2019;648:861-70.
336. Wijkström J, González-Quiroz M, Hernandez M, Trujillo Z, Hultenby K, Ring A, et al. Renal Morphology, Clinical Findings, and Progression Rate in Mesoamerican Nephropathy. American Journal of Kidney Diseases. 2017;69(5):626-36.
337. Elledge M F HRJ, Levine K E, Wickremasinghe R J, Wanigasariya K P, Peiris-John R J. Chronic Kidney Disease of Unknown Etiology in Sri Lanka: Quest for Understanding and Global Implications. (NC) RTP, editor: RTI Press; 2014.
338. Kafle K, Balasubramanya S, Horbulyk T. Prevalence of chronic kidney disease in Sri Lanka: A profile of affected districts reliant on groundwater. Science of The Total Environment. 2019;694:133767.
339. Wimalawansa SJ. Escalating chronic kidney diseases of multi-factorial origin in Sri Lanka: causes, solutions, and recommendations. Environmental health and preventive medicine. 2014;19(6):375-94.

340. Wimalawansa SJ. Public health interventions for chronic diseases: cost–benefit modelizations for eradicating chronic kidney disease of multifactorial origin (CKDmfo/ CKDu) from tropical countries. Heliyon. 2019;5(10):e02309.
341. Liu FX, Rutherford P, Smoyer-Tomic K, Prichard S, Laplante S. A global overview of renal registries: a systematic review. BMC nephrology. 2015;16:31-.
342. Herath C, Jayasumana C, De Silva PMCS, De Silva PHC, Siribaddana S, De Broe ME. Kidney Diseases in Agricultural Communities: A Case Against Heat-Stress Nephropathy. Kidney international reports. 2017;3(2):271-80.
343. Herath N, Wazil A, Kularatne S, Ratnatunga N, Weerakoon K, Badurdeen S, et al. Thrombotic microangiopathy and acute kidney injury in hump-nosed viper (Hypnale species) envenoming: a descriptive study in Sri Lanka. Toxicon : official journal of the International Society on Toxinology. 2012;60(1):61-5.
344. Nanayakkara I, Dissanayake RK, Nanayakkara S. The presence of dehydration in paddy farmers in an area with chronic kidney disease of unknown aetiology. Nephrology. 2020;25(2):156-62.
345. Kaponda P, Muthukrishnan S, Barber R, Holm RH. Drinking water quality and human dimensions of cholera patients to inform evidence-based prevention investment in Karonga District, Malawi. Water Supply. 2019;19(7):2079-87.
346. Msyamboza KP, Kagoli M, M'Bang'ombe M, Chipeta S, Masuku HD. Cholera outbreaks in Malawi in 1998-2012: social and cultural challenges in prevention and control. Journal of infection in developing countries. 2014;8(6):720-6.
347. Mkwate RC, Chidya RCG, Wanda EMM. Assessment of drinking water quality and rural household water treatment in Balaka District, Malawi. Physics and Chemistry of the Earth, Parts A/B/C. 2017;100:353-62.
348. Pritchard M, Mkandawire T, O'Neill JG. Biological, chemical and physical drinking water quality from shallow wells in Malawi: Case study of Blantyre, Chiradzulu and Mulanje. Physics and Chemistry of the Earth, Parts A/B/C. 2007;32(15):1167-77.
349. Edirisinghe EANV, Manthrithilake H, Pitawala HMTGA, Dharmagunawardhane HA, Wijayawardane RL. Geochemical and isotopic evidences from groundwater and surface water for understanding of natural contamination in chronic kidney disease of unknown etiology (CKDu) endemic zones in Sri Lanka. Isotopes in Environmental and Health Studies. 2018;54(3):244-61.
350. Diyabalanage S, Fonseka S, Dasanayake DMSNB, Chandrajith R. Environmental exposures of trace elements assessed using keratinized matrices from patients with chronic kidney diseases of uncertain etiology (CKDu) in Sri Lanka. Journal of Trace Elements in Medicine and Biology. 2017;39:62-70.
351. Anupama YJ, Kiran SK, Hegde SN. Heavy Metals and Pesticides in Chronic Kidney Disease - Results from a Matched Case-Control Study from a Rural Population in Shivamogga District in South India. Indian journal of nephrology. 2019;29(6):402-9.
352. Buser MC, Ingber SZ, Raines N, Fowler DA, Scinicariello F. Urinary and blood cadmium and lead and kidney function: NHANES 2007–2012. International Journal of Hygiene and Environmental Health. 2016;219(3):261-7.
353. Weaver VM, Kotchmar DJ, Fadrowski JJ, Silbergeld EK. Challenges for environmental epidemiology research: are biomarker concentrations altered by kidney function or urine concentration adjustment? Journal of Exposure Science & Environmental Epidemiology. 2016;26(1):1-8.

354. Pearce N, Lawlor DA. Causal inference—so much more than statistics. International Journal of Epidemiology. 2016;45(6):1895-903.
355. Pearce N, Vandenbroucke JP. Causation, mediation and explanation. International journal of epidemiology. 2016;45(6):1915-22.
356. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. Int J Epidemiol. 2016;45(6):1776-86.
357. Pearce N, Vandenbroucke JP, Lawlor DA. Causal Inference in Environmental Epidemiology: Old and New Approaches. Epidemiology. 2019;30(3).
358. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. International Journal of Epidemiology. 2016;45(6):1866-86.
359. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in medicine. 2008;27(8):1133-63.
360. Sekula P, Li Y, Stanescu HC, Wuttke M, Ekici AB, Bockenhauer D, et al. Genetic risk variants for membranous nephropathy: extension of and association with other chronic kidney disease aetiologies. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2017;32(2):325-32.
361. McGinnis DP, Brownstein JS, Patel CJ. Environment-Wide Association Study of Blood Pressure in the National Health and Nutrition Examination Survey (1999–2012). Scientific Reports. 2016;6(1):30373.
362. Sun H, Shi B, Yang F, Wang D. Effects of sulfate on heavy metal release from iron corrosion scales in drinking water distribution system. Water Research. 2017;114:69-77.
363. Sengupta P. Potential health impacts of hard water. International journal of preventive medicine. 2013;4(8):866-75.
364. Dharma-Wardana MWC. Chronic kidney disease of unknown etiology and the effect of multiple-ion interactions. Environmental geochemistry and health. 2018;40(2):705-19.
365. Fianko JR, Laar C, Osei J, Anim AK, Gibrilla A, Adomako D. Evaluation of some heavy metal loading in the Kpeshi lagoon, Ghana. Applied Water Science. 2013;3(1):311-9.

Appendix 1. Administrative boundary levels for Sri Lanka and India

Administrative boundaries are legally documented and attributed jurisdictional boundaries which can range from a group of nation states to small districts and suburbs.

Depending on the country, boundaries are split into varying levels from 0 (entire country) up to 10 (village or ward level). Sri Lanka is divided into nine provinces, which are the first level

administrative division. Districts are the second level division, of which there are 25. Districts are further divided into municipalities which are comprised of Grama Niladhari, or subdivisions, of which there are over 14,000. This is the third-level division. Figure 1 shows these boundary divisions across Sri Lanka.

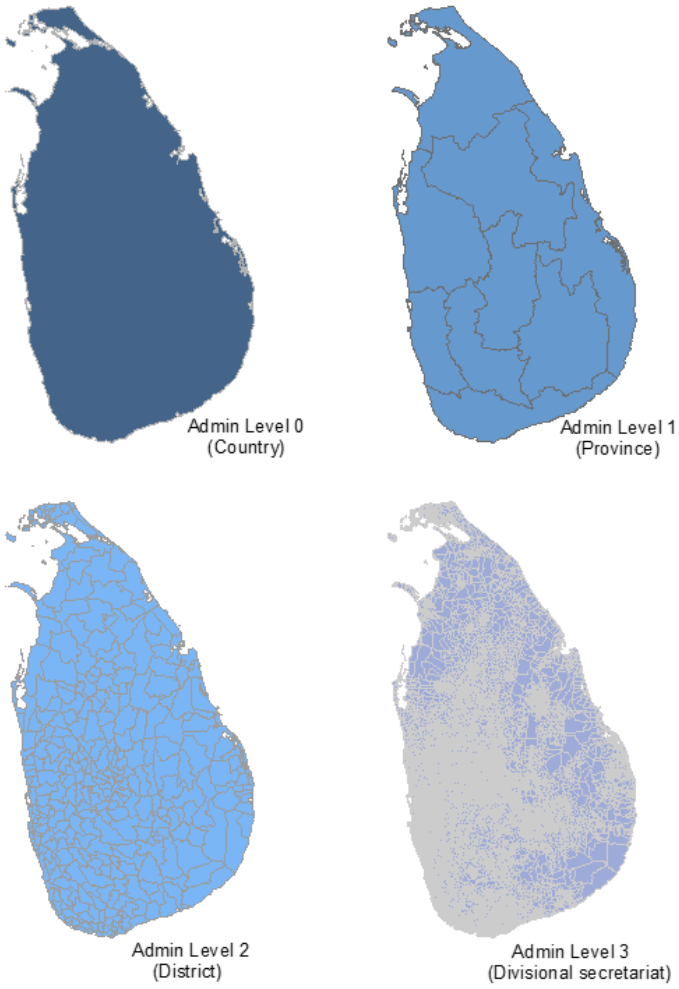


Figure 1. Administrative boundary levels of Sri Lanka

India comprises 28 states and 7 union territories, which are the countries first-level administrative boundaries. The second-level boundaries are districts of which there are 593. The districts are further divided into sub-districts which are known differently in different parts of the country of which there are over 600,000 across India.

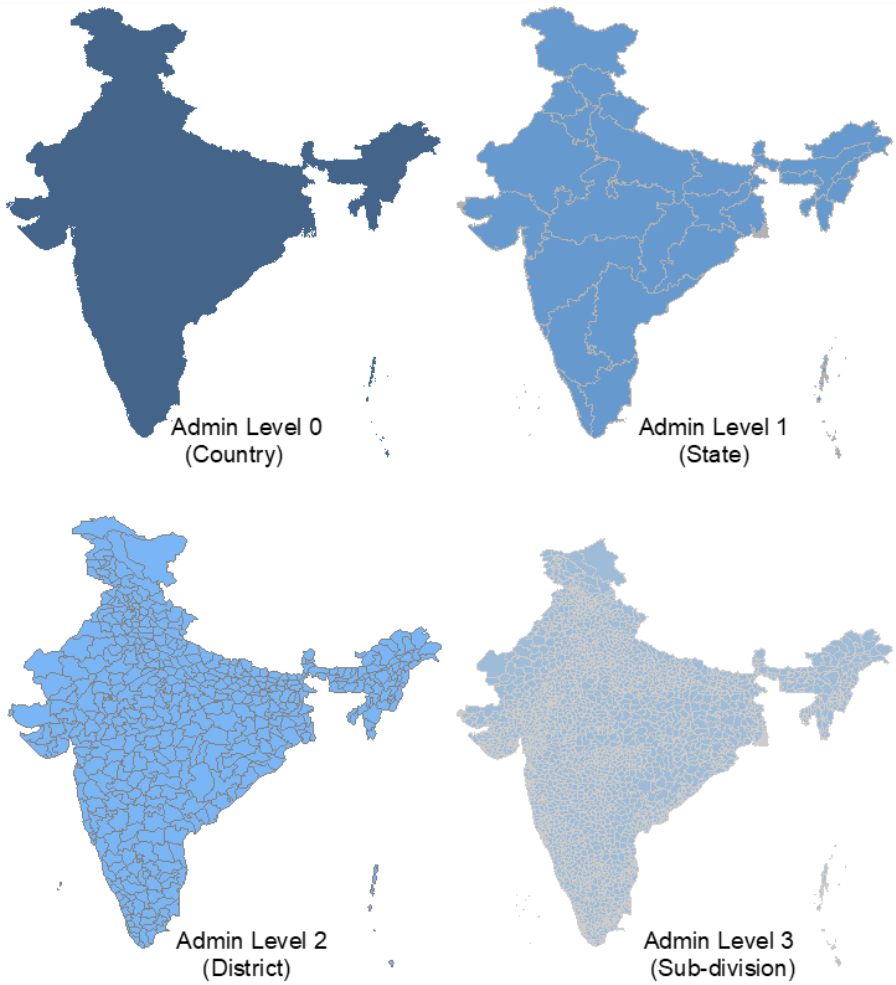


Figure 2. Administrative boundary levels of India [The state of Jammu and Kashmir was not included in any available sub-division level maps]

Appendix 2. Distribution plots for heat index data across Sri Lanka and India

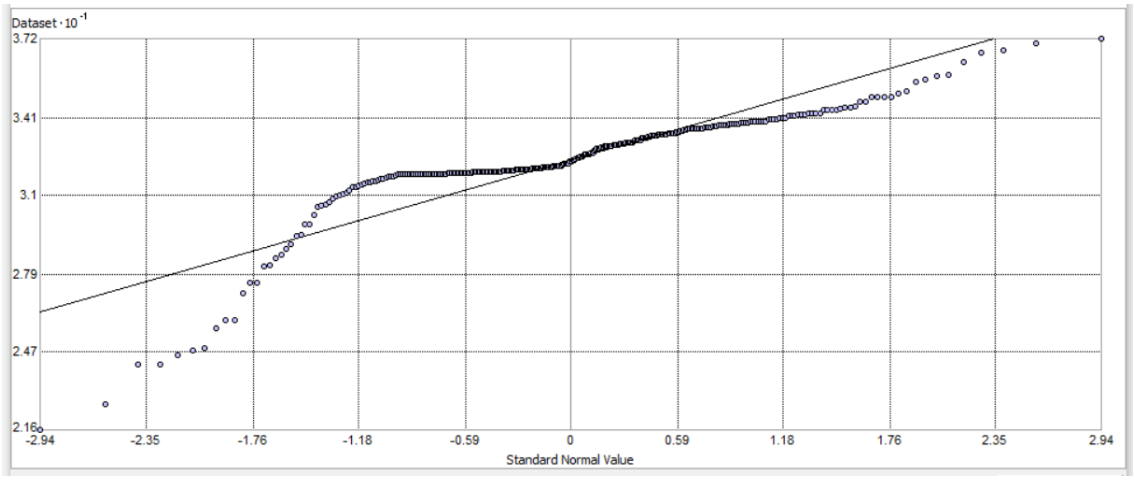


Figure 1. QQ plot of heat index data for Sri Lanka

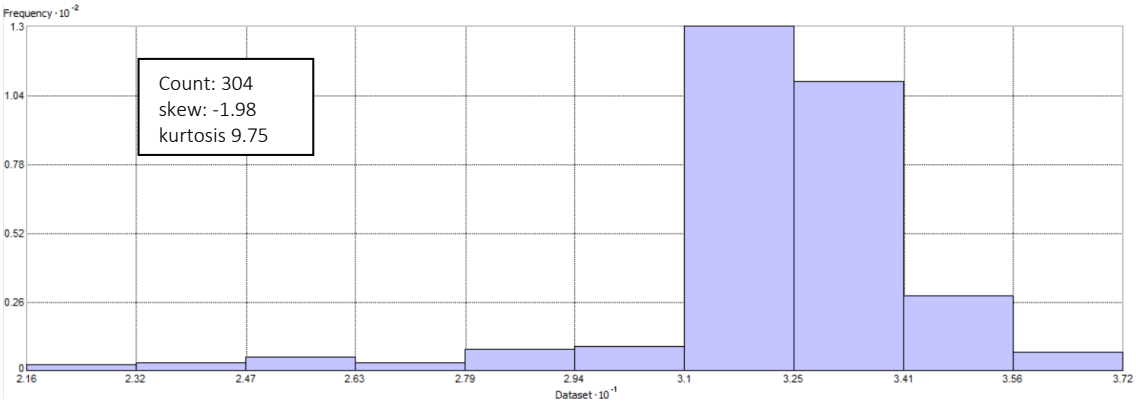


Figure 2. Histogram of heat index data for Sri Lanka

Appendix 3. Cross validation model fit for EBK across India

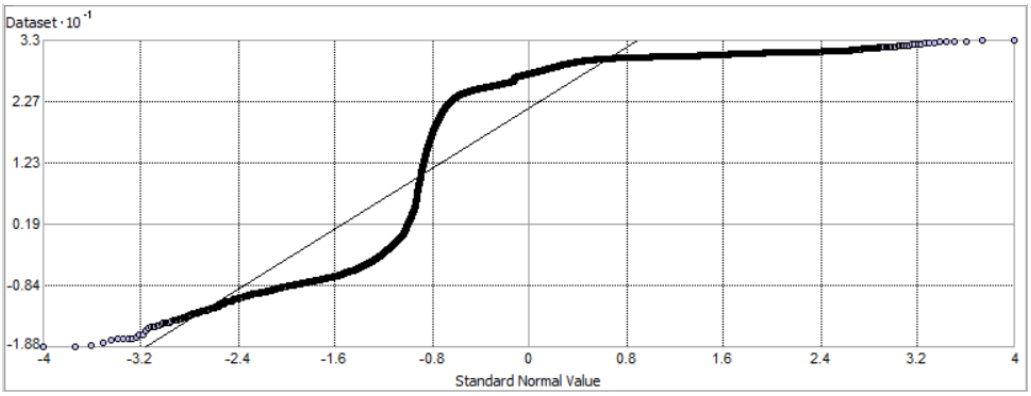


Figure 3. QQ plot of heat index data for the CARRS study period, India

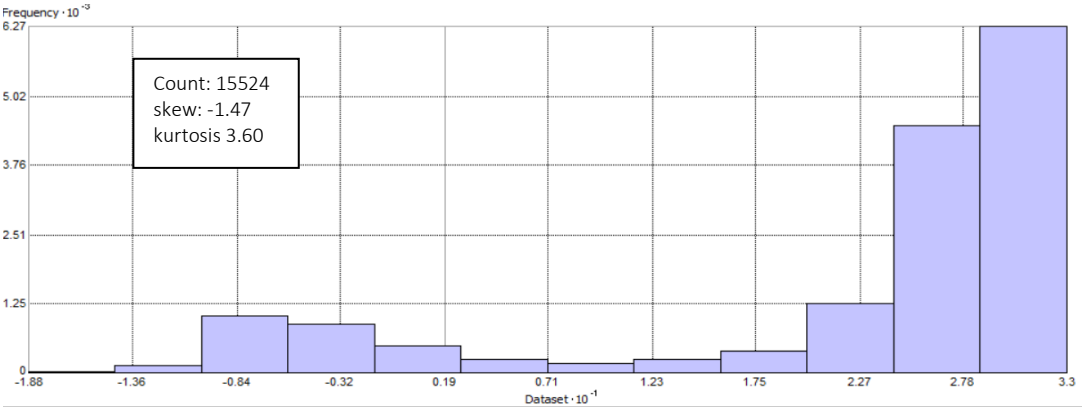


Figure 4. Histogram of heat index data for the CARRS study period, India

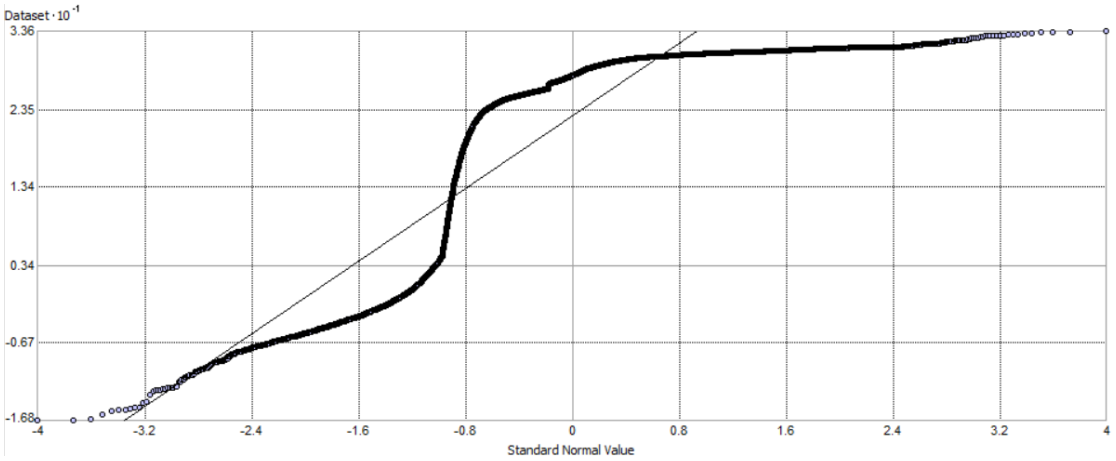


Figure 5. QQ plot of heat index data for the UDAY study period, India

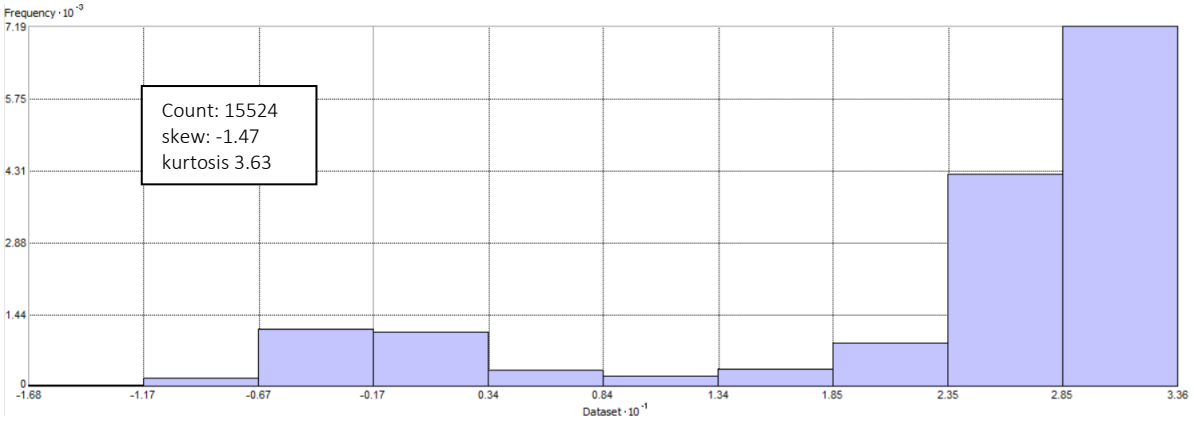


Figure 6. Histogram of heat index data for the UDAY study period, India

Appendix 4. ESA-CCI global and hierarchical land cover trends

LABEL		VALUE		COLOR
GLOBAL	REGIONAL	GLOBAL	REGIONAL	
No Data		0		
Cropland, rainfed		10		
	Cropland, rainfed, herbaceous cover		11	
	Cropland, rainfed, tree or shrub cover		12	
Cropland, irrigated or post-flooding		20		
Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)		30		
Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)		40		
Tree cover, broadleaved, evergreen, closed to open (>15%)		50		
Tree cover, broadleaved, deciduous, closed to open (>15%)		60		
	Tree cover, broadleaved, deciduous, closed (>40%)		61	
	Tree cover, broadleaved, deciduous, open (15-40%)		62	
Tree cover, needleleaved, evergreen, closed to open (>15%)		70		
	Tree cover, needleleaved, evergreen, closed (>40%)		71	
	Tree cover, needleleaved, evergreen, open (15-40%)		72	
Tree cover, needleleaved, deciduous, closed to open (>15%)		80		
	Tree cover, needleleaved, deciduous, closed (>40%)		81	
	Tree cover, needleleaved, deciduous, open (15-40%)		82	
Tree cover, mixed leaf type (broadleaved and needleleaved)		90		
Mosaic tree and shrub (>50%) / herbaceous cover (<50%)		100		
Mosaic herbaceous cover		110		
(>50%) / tree and shrub (<50%)				
Shrubland		120		
	Evergreen shrubland		121	
	Deciduous shrubland		122	
Grassland		130		
Lichens and mosses		140		
Sparse vegetation (tree, shrub, herbaceous cover) (<15%)		150		
	Sparse tree (<15%)		151	
	Sparse shrub (<15%)		152	
	Sparse herbaceous cover (<15%)		153	
Tree cover, flooded, fresh or brakish water		160		
Tree cover, flooded, saline water		170		
Shrub or herbaceous cover, flooded, fresh/saline/brakish water		180		
Urban areas		190		
Bare areas		200		
	Consolidated bare areas	201		
	Unconsolidated bare areas	202		
Water bodies		210		
Permanent snow and ice		220		

Appendix 5, Table 1. Socio-economic and demographic characteristics of study participants in Sri Lanka (n=3296)

Variable	eGFR n= 3297			eGFR categories, n (%)[b] n= 194	n= 3102	
	n (%)[a]	Mean (SD)	p[c]	<60	>60	p[d]
Age (years)			<0.001			<0.001
<20	19 (1)	133.5(20.5)		0	19 (100)	
20-29	549 (17)	114.7(18.2)		1 (1)	548 (99)	
30-39	890 (27)	102.2(17.1)		6 (1)	884 (99)	
40-49	870 (26)	91.3 (18.3)		33 (4)	837 (96)	
50-59	540 (16)	81.8 (17.9)		42 (8)	498 (92)	
60+	428 (13)	69.2 (22.9)		112 (26)	316 (74)	
Sex			<0.001			<0.001
Female	2280 (69)	95.7 (21.7)		83 (4)	2197 (96)	
Male	1016 (31)	90.1 (26.6)		111 (11)	905 (89)	
Zone			0.23			0.01
1	617 (19)	96.5 (22.7)		22 (4)	595 (96)	
2	683 (21)	96.8 24.2)		39 (6)	644 (94)	
3	1996 (61)	92.2 (23.3)		133 (7)	1863 (93)	
Education (years)			<0.001			<0.001
≤5	458 (14)	84.0 (25.5)		60 (13)	398 (87)	
>5≤10	691 (21)	84.0 (23.4)		77 (11)	614 (89)	
>10	2147 (65)	99.3 (21.2)		57 (3)	2090 (97)	
Occupation			0.31			<0.001
Agricultural worker	2312 (70)	88.7 (22.2)		178 (7)	2134 (92)	
Non-agricultural worker	984 (30)	106.2(21.6)		16 (2)	968 (98)	
Household monthly income (RS)			<0.001			<0.001
≤30,000	1779 (54)	90.5 (24.4)		138 (8)	1641 (92)	
>30,000	1474 (45)	98.7 (20.9)		44 (3)	1430 (97)	
Body mass index (kg/m^2)			<0.001			<0.001
Underweight (≤18.5)	601 (18.2)	95.1 (28.0)		54 (9)	547 (91)	
Normal (>18.5 - ≤ 25)	1601 (49)	93.2 (23.2)		97 (6)	1504 (94)	
Overweight (>25 - ≤30)	844 (26)	94.3 (21.2)		37 (4)	807 (96)	
Obese (>30)	250 (8)	95.3 (19.9)		6 (2)	244 (98)	
Fat-free mass (kg)			0.22			0.01
1st tertile (≤37)	2615 (79)	94.9 (23.3)		137 (5)	2478 (95)	
2nd tertile (>37 - <45)	338 (10.3)	90.6 (24.7)		27 (8)	311 (92)	
3rd tertile (≥45)	343 (10.4)	90.6 (22.6)		30 (9)	313 (91)	
Healthy lifestyle choices			<0.001			<0.001
Drinker and/or smoker	822 (25)	86.2 (25.4)		105 (13)	717 (87)	
Non-drinker and/or smoker	2474 (75)	96.6 (22.2)		89 (4)	2385 (96)	
Pesticide user			0.39			<0.001
Yes	1412 (43)	88.4 (22.7)		113 (8)	1299 (92)	
No	1884 (57)	98.1 (23.2)		81 (4)	1803 (96)	
Drinking water source			0.96			0.05
Deep well	2382 (75)	92.8 (23.3)		152 (6)	2230 (94)	
Shallow well	244 (7)	93.4 (23.5)		11 (5)	233 (95)	
Tube well	337 (8)	92.7 (23.2)		22 (7)	315 (93)	
Other	333 (10)	103.5(22.8)		9 (3)	324 (97)	
Heat index (ºC)			0.12			0.59
<33	3180 (96)	94.1 (23.5)		189 (6)	2991 (94)	
>33	116 (4)	91.2 (21.0)		5 (4)	111 (96)	
Altitude (m)			0.10			0.94
<100	1120 (34)	95.8 (24.1)		65 (6)	1055 (94)	
>100	2176 (66)	93.0 (23.1)		129 (6)	2047 (94)	

Land cover			0.80			0.13
Cropland	1203 (36)	95.3 (23.6)		61 (5)	1142 (95)	
Tree/herbaceous cover	1977 (60)	93.2 (23.4)		127 (6)	1850 (94)	
Chena cultivation			0.05			<0.001
Yes	1746 (53)	87.1 (22.8)		157 (9)	1589 (91)	
No	1550 (47)	101.7(22.7)		37 (2)	1513 (98)	
Rice farmer			0.66			<0.001
Yes	2014 (61)	88.1 (22.4)		164 (8)	1850 (92)	
No	1282 (39)	103.2(22.1)		30 (2)	1252 (98)	
Vegetable farmer			0.05			0.05
Yes	704 (21)	87.9 (22.0)		54 (8)	650 (92)	
No	2592 (79)	95.6 (23.6)		140 (5)	2452 (95)	

[a] percentage in columns; [b] percentage in rows; [c] Bartlett's test for equal variance; [d] Chi-square test

Appendix 6, Table 1. Socio-economic and demographic characteristics of study participants in India (n=11,119)

[a] percentage in columns; [b] percentage in rows; [c] Bartlett's test for equal variance; [d] Chi-square test

Variable	eGFR n= 11,119			eGFR categories, n (%)[b] n= 165	n= 10954	
	n (%)[a]	Mean (SD)	p[c]	<60	>60	p[d]
Age (years)			<0.001			<0.001
20-29	1079 (10)	125.5 (11.4)		1 (1)	1078 (99)	
30-39	4076 (37)	112.8 (13.1)		12 (1)	4064 (99)	
40-49	3223 (29)	103.8 (13.7)		33 (2)	3190 (98)	
50-59	1581 (14)	95.0 (14.1)		32 (4)	1549 (96)	
60+	1160 (10)	84.3(15.9)		87 (10)	1073 (90)	
Sex			<0.001			<0.001
Female	6423 (58)	108.8 (16.8)		67 (1)	6356 (99)	
Male	4696 (42)	102.1 (17.9)		98 (2)	4598 (98)	
Area			<0.001			<0.001
Urban	7663 (69)	109.1 (16.4)		56 (1)	7607 (99)	
Rural	3456 (31)	99.0 (17.9)		109 (3)	3347 (97)	
Latitude			<0.001			0.05
North	4978 (45)	104.9 (16.8)		59 (1)	4919 (99)	
South	6141 (55)	106.9 (18.1)		106 (2)	6035 (98)	
Education (years)			<0.001			<0.001
≤5	3366 (14)	101.5 (18.7)		108 (3)	3258 (97)	
>5≤10	4584 (21)	108.8 (16.5)		23 (1)	4561 (99)	
>10	3169 (65)	107.5 (16.7)		34 (1)	3135 (99)	
Occupation			<0.001			<0.001
Employed	5581 (53)	104.8 (16.7))		77 (1)	5504 (99)	
Unemployed	5538 (47)	107.3 (18.3)		88 (1)	5450 (99)	
Household monthly income (RS)			0.15			<0.001
≤30,000	10,000 (89)	106.6 (17.5)		149 (2)	9862 (98)	
>30,000	831 (7)	100.9 (16.8)		10 (1)	821 (99)	
Unknown	277 (4)	99.7 (18.4)		6 (2)	271 (98)	
Body mass index (kg/m²)			<0.001			<0.001
Underweight (≤18.5)	1442 (13)	104.8 (19.3)		36 (2)	1465 (98)	
Normal (>18.5 - ≤ 25)	5191 (47)	105.1 (17.8)		99 (2)	5092 (98)	
Overweight (>25 - ≤30)	2921 (26)	106.3 (16.4)		22 (1)	2897 (99)	
Obese (>30)	999 (9)	109.2 (16.5)		4 (1)	997 (99)	
No data	566 (5)	0		4 (1)	562 (99)	
Fat-free mass (kg)			<0.001			0.08
1st tertile (≤37)	3830 (34)	106.8 (18.0)		71 (2)	3759 (98)	
2nd tertile (>37 - <45)	3691 (33)	106.5 (17.1)		54 (1)	3637 (99)	
3rd tertile (≥45)	3002 (27)	102.7 (16.8)		36 (1)	2966 (99)	
No data	597 (5)	0		4 (1)	593 (99)	
Healthy lifestyle choices			<0.001			<0.001
Drinker and/or smoker	3378 (34)	101.0 (18.6)		81 (2)	3297 (98)	
Non-drinker and/or smoker	7741 (66)	108.1 (17.2)		84 (1)	7657 (99)	
Vegetarian			<0.001			0.38
Yes	3686 (33)	102.8 (16.3)		49 (2)	3637 (98)	
No	7433 (67)	107.5 (18.0)		116 (1)	7317 (99)	
Heat index (°C)			0.55			0.22
<28	5984 (54)	103.6 (17.31)		97 (2)	5887 (98)	
>28	5135 (46)	108.7 (17.44)		68 (1)	5067 (99)	
Altitude (m)			0.001			0.80
<100	5788 (52)	107.54 (17.8)		88 (2)	5700 (98)	
>100	5331 (48)	104.26 (17.1)		77 (1)	5254 (99)	
Land cover			<0.001			<0.001
Cropland	4951 (45)	101.6 (17.98)		114 (2)	4837 (98)	
Urban	6168 (55)	109.5 (16.39)		51 (1)	6117 (99)	

Appendix 7. Cross validation model fit for EBK across Sri Lanka

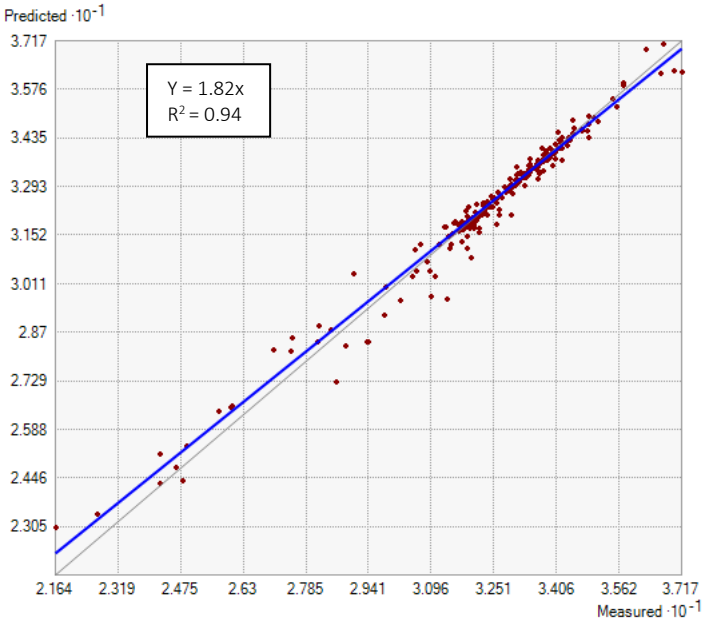


Figure 1. Cross validation model fit for EBK interpolation of back-transformed data, Sri Lanka

Appendix 8. Validation model fit for interpolation methods across Sri Lanka for sensitivity analyses

Figure 1 (a) Histogram of heat index data when restricted to altitudes within 100m of the study zones

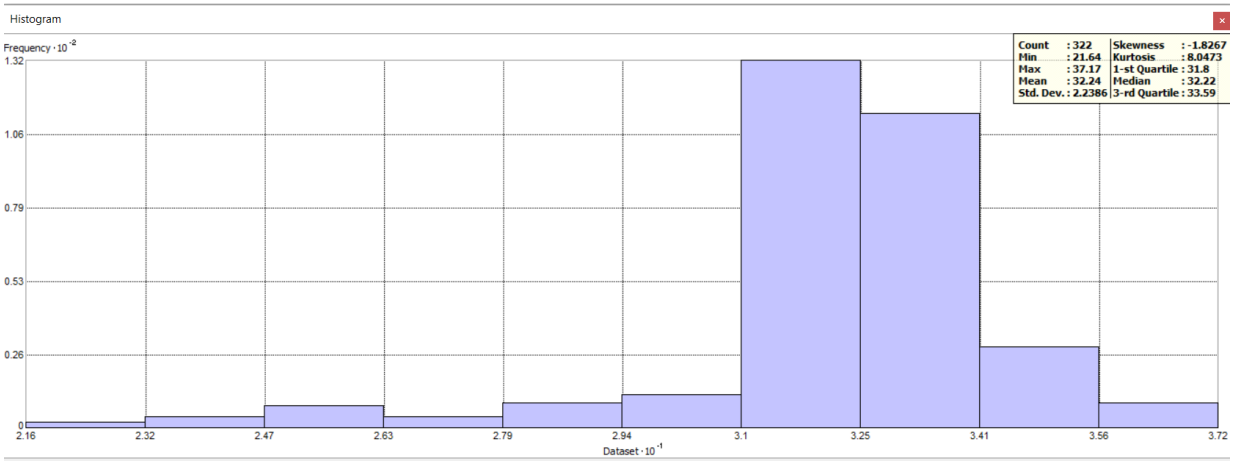


Figure 1 (a) Histogram of heat index data when restricted to altitudes within 100m of the study zones

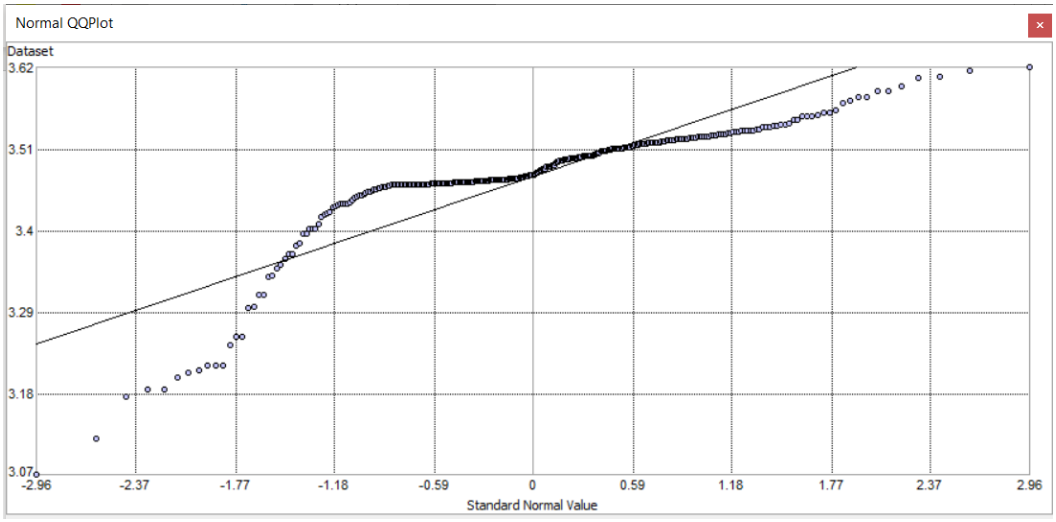


Figure 1 (b) QQ plot of heat index data when restricted to altitudes within 100m of the study zones

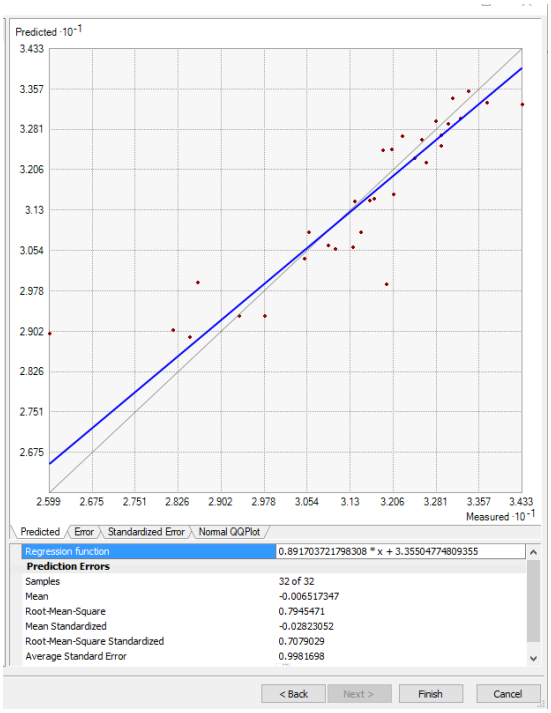


Figure 2. Model fit for kriging method across Sri Lanka restricted altitude sample

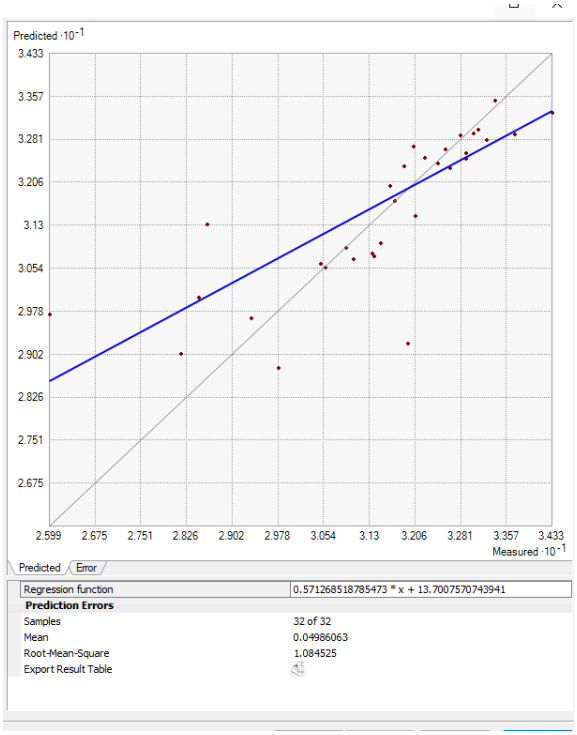


Figure 3. Model fit for inverse distance weighting method across Sri Lanka restricted altitude sample

Appendix 9. Sensitivity analysis assessing effects of work group and chena cultivation

Table 1. Linear regression sensitivity analysis investigating the effect of variables work group and chena cultivation on each other, and also include bilinear interpolation heat index data

Variable	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]
	eGFR Coefficient (95%CI)			
Age [e]	-10.78***	-9.96***	-9.95***	-9.92***
(10-year increase)	(-11.26, -10.34)	(-10.54, -9.37)	(-10.54, -9.36)	(-10.51, -9.33)
Sex [f]				
Male	-1.16	7.81***	7.84***	7.78***
	(-2.53, 0.20)	(5.26, 10.36)	(5.30, 10.39)	(5.24, 10.33)
Female	REF	REF	REF	REF
Education (years)				
≤5	0.50	0.19	0.30	0.31
	(-1.49, 2.50)	(-1.82, 2.25)	(-1.73, 2.35)	(-1.72, 2.16)
>5≤10	-2.24**	-1.60	-1.44	-1.46
	(-4.88, -0.55)	(-3.30, 0.12)	(-3.17, 0.27)	(-3.45, 0.25)
>10	REF	REF	REF	REF
Pesticide user				
Yes	-2.91***	-1.08	-1.14	-0.95
	(-4.30, -1.52)	(-2.63, 0.50)	(-2.68, 0.40)	(-2.53, 0.61)
No	REF	REF	REF	REF
Household monthly income (RS) [g]				
≤30,000	-0.73	-0.72	-0.66	-0.64
	(-2.04, 0.57)	(-2.06, 0.61)	(-2.00, 0.68)	(-1.98, 0.70)
>30,000	REF	REF	REF	REF
Healthy lifestyle choices				
Non-drinker and/or smoker	REF	REF	REF	REF
Drinker and/or smoker	-8.49***	-7.33***	-7.39***	-7.37***
	(-10.88, -6.11)	(-9.74, -4.88)	(-9.82, -4.97)	(-9.80, -4.94)
BMI (kg/m^2)	-0.39	-0.35	-0.31	-0.34
5kg/m^2 increase	(-1.04, 0.25)	(-0.47, 0.98)	(-0.40, 1.04)	(-1.01, 1.06)
Fat Free Mass (kg)	-0.90***	-0.71	-0.71	-0.71
5 kg/m2 increase	(-1.23, -0.59)	(-0.38, 1.07)	(-1.08, -0.35)	(-1.09, -0.35)
Water source				
Deep well	-1.42	-1.55	-1.67	-1.66
	(-3.83, 0.98)	(-3.96, 0.79)	(-4.05, 0.70)	(-4.53, 1.44)
Shallow well	REF	REF	REF	REF
Tube well	-1.15	-1.47	-1.57	-1.54
	(-4.16, 1.85)	(-4.52, 1.45)	(-4.56, 1.41)	(-2.27, 1.44)
Other source	1.69	1.13	1.15	1.80
	(-1.34, 4.73)	(-1.92, 4.10)	(-1.85, 4.16)	(-1.92, 4.10)
Chena cultivation				
Yes	-2.43***		-1.58*	-1.16
	(-3.82, -1.06)		(-3.25, -0.07)	(-1.60, 1.22)
No	REF		REF	REF
Rice farmer				
Yes	-2.21 **	0.67	-1.11	076
	(-3.66, -0.76)	(-1.59, 2.93)	(-1.87, 1.63)	(-3.40, 3.04)
No	REF	REF	REF	REF
Vegetable farmer				

Yes	-2.93**	-1.77*	-2.10*	-1.88*
	(-4.47, -1.41)	(-3.43, -0.19)	(-3.69, -0.52)	(-3.51, -0.26)
No	REF	REF	REF	REF
Land cover				
Cropland	0.99*	0.65	0.61	0.62
	(-0.29, 2.26)	(-0.64, 1.95)	(-0.68, 1.91)	(-0.67, 1.92)
Tree/herbaceous cover	REF	REF	REF	
Altitude				
50m increment	-3.03**	-5.02*	-5.92***	-5.91***
	(-4.47, -1.60)	(-9.04, -3.01)	(-8.97, -2.88)	(-8.96, -2.87)
Heat index EBK (ºC)				
0.5 ºC increment	-3.03**	-2.84***	-2.91***	-3.89***
	(-4.47, -1.60)	(-4.94, 0.17)	(-4.38, -1.45)	(-4.36, -1.43)
Occupation				
Agricultural worker	-3.34***	-2.37**		-1.70
	(-4.88, -1.80)	(-4.94, -0.17)		(-4.47, 1.06)
Non-agricultural worker	REF	REF		REF
Chena cultivation				
Yes	-2.43***		-1.58*	-1.16
	(-3.82, -1.06)		(-3.25, -0.07)	(-1.60, 1.22)
No	REF		REF	REF
Heat index Bilinear interpolation (ºC)[h]	-6.20**	-5.87***	-5.00***	-4.40***
0.5ºC increment	(-8.21, -4.21)	(-7.97, -3.78)	(-8.19, -3.96)	(-6.61, -2.21)

[a] minimal adjustment for age, sex , [b] variables mutually adjusted for age, sex, education, income, healthy lifestyle choices, BMI, FFM, water sources, farming, EBK heat index and DEM chena cultivation omitted; [c] Variables adjusted as before in the absence of work group[d] All variable mutually adjusted; [e] adjusted for sex [f] adjusted for age, [g] Exchange rate (RS to USD) 0.001 at time of questionnaire; [h]Model adjustment for bilinear interpolation heat index data as comparison to EBK. Coefficients remained largely unchanged and are appended; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol

Appendix 10. Validation model fit for EBK for the CARRS and UDAY study periods, India

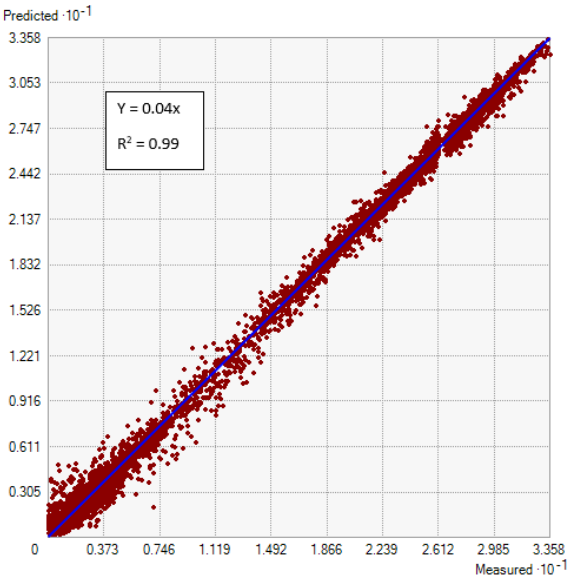


Figure 1. Cross validation model fit for EBK interpolation, CARRS India

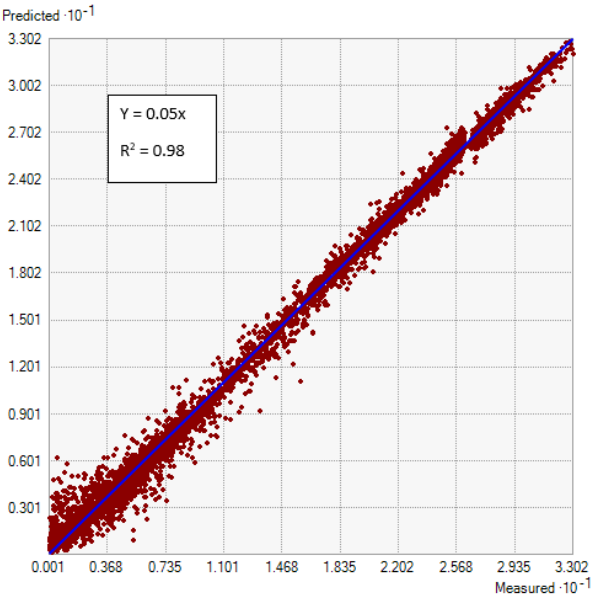


Figure 2. Cross validation model fit for EBK interpolation, UDAY India

Appendix 11. Model fit for interpolation methods across India as part of sensitivity analyses

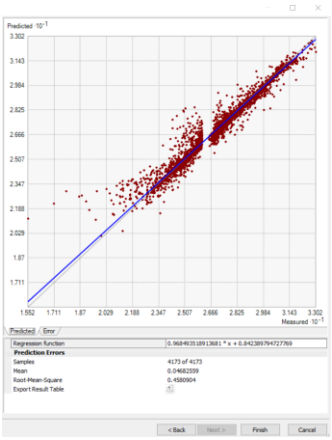


Figure 1. Model fit for inverse distance weighting

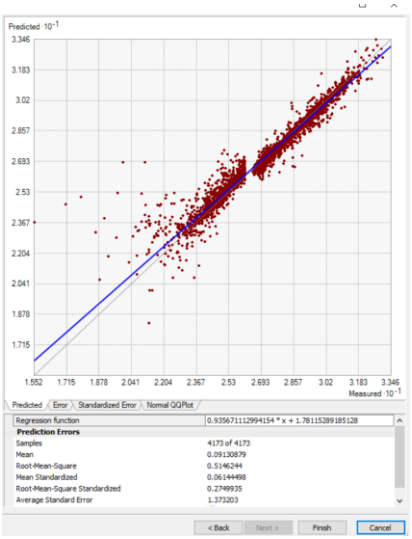


Figure 2. Model fit for kriging across India

Appendix 11. Ethical approval documents (a) Malawi National Health Sciences Research Committee protocol number #1912 (Lilongwe, Malawi) (b) London School of Hygiene and Tropical Medicine Ethics Committee protocol #14709 (London, UK)

(a)

Telephone: + 265 789 400
Facsimile: + 265 789 431

All Communications should be addressed to:

The Secretary for Health and Population

In reply please quote No.

.....

MINISTRY OF HEALTH AND POPULATION
P.O. BOX 30377
LILONGWE 3
MALAWI

25th January, 2018

Amelia Crampin
MEIRU
Lilongwe.

Dear Sir/Madam,

RE: Protocol 17/09/1912: Cross Sectional Study on Prevalence and Risk Factors of Chronic Kidney Disease of Unknown Cause (CKDu) in Malawi Comparing Urban and Rural Populations.

Thank you for the above titled proposal that you submitted to the National Health Sciences Research Committee (NHSRC) for review. Please be advised that the NHSRC has **reviewed** and **approved** your application to **conduct** the above titled study.

- **APPROVAL NUMBER : 1912**
- The above details should be used on all correspondences, consent forms and documents as appropriate.
- **APPROVAL DATE : 25/01/2018**
- **EXPIRATION DATE**
  This approval expires on **24/01/2019**. After this date, this project may only continue upon renewal. For purposes of renewal, a progress report on a standard form obtainable from the NHSRC Secretariat should be submitted one month before the expiration date for continuing review.
- **SERIOUS ADVERSE EVENT REPORTING:** All serious problems having to do with subject safety must be reported to the NHSRC within 10 working days using standard forms obtainable from the NHSRC Secretariat.
- **MODIFICATIONS**: Prior NHSRC approval using forms obtainable from the NHSRC Secretariat is required before implementing any changes in the protocol (including changes in the consent documents). You may not use any other consent documents besides those approved by the NHSRC.
- **TERMINATION OF STUDY**: On termination of a study, a report has to be submitted to the NHSRC using standard forms obtainable from the NHSRC Secretariat.
- **QUESTIONS**: Please contact the NHSRC on phone number *+265 888 344 443* or by email on mohdoccentre@gmail.com.
- **OTHER**: Please be reminded to send in copies of your final research results for our records (Health Research Database).

SECRETARY FOR HEALTH
2018 -01- 25
P.O BOX 30377, CAPITAL CITY, LILONGWE 3

Kind regards from the NHSRC Secretariat.

.....

For: **CHAIRPERSON, NATIONAL HEALTH SCIENCES RESEARCH COMMITTEE**
Promoting Ethical Conduct of Research[1]

Executive Committee: *Dr B. Chilima (Chairperson), Dr B. Ngwira (Vice-Chairperson)*
Registered with the USA Office for Human Research Protections (OHRP) as an International IRBIRB Number IRB00003905 FWA00005976

11(b)

London School of Hygiene and Tropical Medicine

Keppel Street, London WC1E 7HT

United Kingdom

Switchboard: +44 (0)20 76368686

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Professor Neil Pearce
Professor
Department of Medical Statistics (MSD)
Epidemiology and Population Health (EPH)
LSHTM
10 May 2018

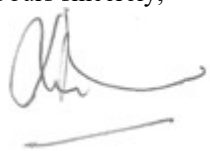
Dear Neil,

Study Title: Chronic Kidney Disease of unknown cause (CKDu) in disadvantaged communities in low-and-middle income countries (LMICs): a cross-sectional prevalence and risk factor study in Malawi
LSHTM Ethics Ref: 14709
Thank you for responding to the Observational Committee's request for further information on the above research and submitting revised documentation.
The further information has been considered on behalf of the Committee by the Chair.
Confirmation of ethical opinion
On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.
Conditions of the favourable opinion
Approval is dependent on local ethical approval having been received, where relevant.
Approved documents
The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Local Approval	CKDu NHSRC submission 1	30/08/2017	1.0
Local Approval	Malawi ethics letter 1	06/10/2017	1.0
Local Approval	Malawi ethics letter 2	30/10/2017	1.0
Local Approval	NHSRC resubmission 1	30/10/2017	1.0
Local Approval	Malawi ethics letter 3	23/11/2017	1.0
Local Approval	NHSRC resubmission 2	23/11/2017	1.0
Local Approval	Malawi ethics letter 4	30/11/2017	1.0
Investigator CV	Neil Pearce 1pg CV	04/01/2018	1.0
Information Sheet	Malawi information sheet	04/01/2018	1.0
Information Sheet	Malawi consent form	04/01/2018	1.0
Information Sheet	Malawi consent form – local language	04/01/2018	1.0
Investigator CV	Mia Crampin 1pg CV	23/01/2018	1.0

Protocol / Proposal	CKDu NHSRC study protocol	23/01/2018	1.0
Investigator CV	Moffat Nyirenda 1pg CV	23/01/2018	1.0
Investigator CV	Wisdom Nkanga 1pg CV	23/01/2018	1.0
Investigator CV	Ben Caplin 1pg CV	23/01/2018	1.0
Investigator CV	Liam Smeeth 1pg CV	23/01/2018	1.0
Advertisements	CKDu NHSRC recruitment	24/01/2018	1.0
Protocol / Proposal	ETE 2016 NCD methods	24/01/2018	1.0
Protocol / Proposal	IJE 2012 DSS profile	24/01/2018	1.0
Covering Letter	2018 04 1 CKD Malawi ethics response	25/04/2018	1
Covering Letter	2018 04 3 Malawi local ethics final approval	25/04/2018	1

After ethical review
The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review
using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.
The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project
by submitting a Serious Adverse Event form.
An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.
At the end of the study, the CI or delegate must notify the committee using an End of Study form.
All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk
Additional information is available at: www.lshtm.ac.uk/ethics
Yours sincerely,



Professor John DH Porter
Chair
ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/
Improving health worldwide

Appendix 12: DEGREE information sheet English and Chichewa (Area 25)

MALAWI EPIDEMIOLOGY AND INTERVENTION RESEARCH UNIT POLOGALAMU YA LSHTM YA KAFUKUFUKU WA MATENDA OSAPATSIRANA

CHRONIC KIDNEY DISEASE OF UNKNWON ORIGIN STUDY INFORMATION SHEET

To be read out to potential participants of the **CKDu study in Lilongwe.**

PARTICIPANT INFORMATION SHEET

Study title: Chronic Kidney Disease of unknown cause (CKDu) in disadvantaged communities in low-and-middle income countries

(LMICs)

Please read this leaflet carefully. It explains the research study we are currently conducting and ask for your help in participation. Please ask the researcher who explained the study if you require any further information. Taking part is purely voluntary and whether or not you agree to join the study will not change the care you receive in any way.

1. **Why is the study being done?**
   Chifukwa chiyani kafukufuku uyu akuchitika?

   Researchers at the Malawi Epidemiology and Intervention Research Unit (MEIRU) and the London School of Hygiene & Tropical Medicine are interested in a disease that affects kidneys in a long term called Chronic Kidney Disease of Unknown Origins (CKDu) and you are being invited to participate. This is a condition where a person's kidneys begin to stop working properly, but we cannot figure out why they are failing. We want to find out how common these diseases are, and why people develop them. We would like to understand more about the severity of these problems among Malawians.

   Opanga kafukufuku aku Malawi Epidemiology and Intervention Research Unit (MEIRU) ndi sukulu ya u kachenjede ya London School of Hygiene and Tropical Medicine akupanga kafukufuku wa matenda a

impsyo opanda chifukwa choyambila omwe akuchitika kuno ku Malawi komanso mayiko ena a dziko la pansi ndipo mukuyitanidwa kutenga nawo gawo. Tikufuna kufufuza kuti matendawa ndi ochuluka bwanji komanso chifukwa chiyani anthu ambiri akupezeka ndi nthendayi.

2. Who is funding the study and has the study been approved ethically?
Kafukufuku atenga nthawi yaitali bwanji ndipo ndalama zichokela kuti ndipo a zaumoyo a loleza kuti achitike?

The research study will last for two years with funding being provided by the Medical research council, a major UK based research charity. The study proposal has been approved by National Health Sciences Research Committee in Malawi and by the London School of Hygiene & Tropical Medicine Ethics committees.

Kafukufukuyu ndi wa zaka ziwili ndipo akupeleka chithandizo ndi a Medical Research Council. Kafukufuku ameneyi walandila chilolezo kuchokela kwa National Health Sciences Research Committee amene amayang'anila 375ioim za kafukufuku muno mu Malawi.

3. Why have I been asked to take part?
Nchifukwa chiyani ndapemphedwa kutengapo mbali?

We are inviting people from Area 25 in Lilongwe and Chilumba in Karonga to take part in the study. Kindly read this information sheet about the study and ask any questions you have.

Tikuitana anthu kuchokera ku dera la 25, Lilongwe ndi Chilumba, Karonga kutengapo mbali pa kafukufukuyi. Chonde werengani uthenga omwe uli pa pepalali okhuza za kafukufukuyu ndi kufunsa mafunso ngati muli nawo.

4. Is taking part in the study voluntary?
Kutengapo mbali mukafukufukuyi ndi kongothandiza?

It is entirely your decision to take part in the study. You can discontinue at any point of time. Your medical care will not be affected by your decision.

Mukhonza kusankha kutenga nawo mbali mukafukufukuyu. Tidzakulongosoleraniza kafukufukuyu ndikukuyankhani mafunso onse amene mungakhale nawo muchikalatachi. Komabe muli ndi ufulu kusiya kutenga nawo mbali nthawi iliyonse popanda kupeleka chifukwa ndipo izi sizizasokoneza chithandizo cha mankhwala ena alionse omwe mumalandila ku chipatala.

5. What does it mean to participate?
Zikutanthauza chani kutengapo mbali?

We will ask you some questions about your health and lifestyle (for example your tobacco and alcohol consumption, physical activity, your diet). We will measure your weight, height, waist and hip size, blood pressure and pulse. The study team will come to your house in the morning while you fasted overnight – that is not to eat or drink anything. You will be asked to give a small amount of blood. You can choose to give blood or refuse if you wish. If you agree, we would like to draw a small amount (18 ml – about a tablespoon). We will fit the blood into 3 small tubes. Trained personnel will draw the blood. The supplies used for drawing blood will be safe and sterile and used only once and the supplies will be destroyed after use. A small amount of blood will be used to conduct a rapid HIV test. A small amount of the blood will be used to test for aspects of renal function. After this we will ask you to submit a small amount of urine (30 ml – about a small Vaseline bottle) that will be used to see how the kidneys are functioning.

Tikufunsani mafunso okhudza nthanzi ndi moyo wanu chitsanzo (kasutidwe ka fodya ndi kamwedwe ka mowa, kulimba thupi ndi kadyedwe) Tikuyezani kulemera, msinkhu, chiuno, kuthamanga kwa magazi ndi kugunda kwa mtima. Akafukufuku tizambwela kwanu m'mamawa musanadye kanthu kuchoka madzulo, pamenepo ndiye kuti simunadye kapena kumwa kanthu. Muzapemphedwa kupereka magazi pang'ono. Mutha kusankha kupereka magazi kapena ayi ngati musakufuna. Ngati mutavomera tidzafuna kutengapo magazi ma tube ang'onoang'ono atatu. Akadaula ndiwo adzatenge magaziwa. Zipangizo zogwiritsa ntchito kutenga magazi zidzakhala zotetezedwa zomwe potha pa ntchitoyi zidzatayidwa. Mulingo ochepa wa magazi udzagwiritsidwa ntchito poyeza HIV. Mulingo ochepa wa magazi udzagwiritsidwa ntchito poyesa zakagwiridwe ntchito ka impsyo.Pakutha pa izi tizakupemphani kupereka mikodzo yochepa (30ml, kapena mulingo wa kabotolo kakang'ono ka Vaseline)

6. Why is blood being collected in this study and what will happen to the samples.
Nchifukwa chiyani magazi akugwiritsidwa ntchito mukafukufukuyi, chidzachitike ndi chani ku zoteleledwa?

Chronic Kidney Disease affects how well your kidneys work. We will test how well your kidneys are filtering the blood. We will also look for diseases such as HIV. Additionally, we would like to store some of the blood to test for biomarkers linked to NCDs. The blood you give will be used for research purposes only. It will be taken to our laboratory for testing. All results will be kept confidential.

Matenda a Impsyo amakhudza m'mene Impsyo zanu zimagwirira ntchito. Tidzayesa m'mene kagwiridwe ka ntchito ka Impsyo zanu kakuyendera posefa magazi. Tidzayan'gananso matenda ena monga HIV. Kuphatikiza apo, tidzafuna kusunga magazi kuti tidzayese zoyambitsa matenda osapatsirana. Magazi omwe mudzapereke cholinga chake nkugwiritsa ntchito mukafukufukuyi basi. Adzatengedwera 376ioimp oyesera. Zotsatira 376ioim zizasungidwa mwachinsisi.

7. Can you find out what our measurements show and the results of the blood tests.
Mungapeze zomwe mayesedwe athu akuonetsa ndi zotsatira za kuyesa kwa magazi?

We can tell you about physical measurements (e.g. your weight), and your blood pressure straight away. If you want us to, we can contact you if anything abnormal is suggested in your blood results. We can phone you. The study staff will explain what your results mean. They will advise you on whether you need to seek further testing at the area25 clinic, your local clinic or hospital. If you are referred to a Hospital or a clinic, the medical staff there may recommend changes to your lifestyle or treatment to help keep you healthy. If your blood results are normal, you will not be contacted, but you can still contact us to confirm.

Titha kukuwuzani za kayesedwe kena monga kulemera, kuthamanga kwa magazi pompo. Ngati mungafune titha kulumizana nanu ngati zina zachilendo zingapezedwe nkuyesedwa kwa magazi anu. Titha kukuyimbirani phone. Ogwira ntchito a kafukufuku adzakufotokozerani ngati padzafunike zoyesa zina pa chipatala chaching'ono cha m'dera la 25, kapena chipatala chachikulu. Ngati muzauzidwe kupita ku chipatala chachikulu kapena chaching'ono, ogwira ntchito zachipatala atha kuzakulangizani kusintha kakhalidwe ka moyo wanu kapena kupereka thandizo lamankhwala kuthandizira thanzi lanu. Ngati zotsatira za magazi anu zidzakhale bwino, muzalumikizidwa komanso mutha kutifunsa pofuna kungotsimikiza.

8. Can I decline to be in the study or decline to give blood?
Kodi nditha kukana kukhala nawo mukafukufuku kapena kukana kupereka magazi?

It is up to you whether or not to take part in the study. You can stop being in the study or decide not to give blood at any time you want without giving a reason. This will not affect your standard medical care.

Awa ndi maganizo anu kusankha kutenga nawo mbali mukafukufukuyu, mutha kusiya kutenga mbali

mukafukufukuyu kapena kusankha kusapereka magazi nthawi iliyonse popanda kupereka chifukwa. Izi sizidzakhudza za chithandizo chanu kachuchipatala.

9. What are the benefits from being in the study?
Kodi phindu lotenga nawo gawo mukafukufuyu ndi chani?

If you have abnormal results, we will refer you to your clinic for further tests, where you may receive lifestyle advice or treatment to improve your health. By taking part in this study you will help researchers gain a clearer understanding of Chronic Kidney Disease of Unknown origin. Your participation will also help them understand some of the causes of this disease and why there are differences between rural and urban populations. This would further help in prevention and management of this disease.

Ngati zotsatira zakuyesedwa kwanu sizizakhala zabwino tidzakutumizani kuchipatala kuti akapitirize kuyesa zina. Pamene muzapatsidwe uphungu wa kakhalidwe kapena thandizo lamankhwala kupititsa patsogolo thanzi lanu. Kutengapo mbali pakafukufukuyu muzamuthandiza wakafukufukuyo kumvetsa za matenda a Impsyo ndi zoyambitsa zake zomwe sizidziwika. Kutengapo mbali kwanunso kuzathandiza wakafukufuku kudziwa zina zoyambitsa matendawa ndi chifukwa chiyani pamakhala kusiyana pa kuchuluka kwa anthu okhala kumudzi ndi ku town. Izi zidzapitirira kuthandiza kupewa ndi kusamalira nthendayi.

10. What are the risks of being in the study?
Kodi kuopsa kotengapo mbali mukafukufukuyi ndi chani?

We do not expect that you will incur any risks by participating in this study. Blood drawing may cause a small amount of discomfort when the needle is passed into the vein but does not cause long term pain. After blood is taken sometimes there can be a small bruise or soreness at the site. You will not incur any costs as a result of your participation in this study.

Mwina mutha kumva kupweteka pang'ono pamalo amene mwabayidwa potenga magazi.Komabe zisakupangitseni kukhala ndi nkhawa chifukwa amene azizakutengani magaziwo ndi ophunzitsidwa bwino pa katengedwe ka magazi. Zipangizo zomwe timagwilitsa ntchito potenga magazi ngati singano ndi zotetezedwa bwino.

11. How will the information blood and urine samples I give in the study be kept private?
Kodi uthenga uliwonse okhudza mikodzo komanso magazi anga zidzatetezedwa bwanji?

All information which is collected about you in this study will be kept strictly confidential. To protect your privacy, we will use a unique code number to identify you and all information about you, including your blood and urine samples. Under no circumstances will your name, address or telephone number be given to any researchers. We will keep your records and samples securely locked. For the current study, only research team members from this study will be able to look at the data and samples. Your name or any other facts that might point to you will not appear when we present this study or publish its results.

Uthenga wina uliwonse okhudza inu mukafukufukuyu udzasungidwa mwachinsisi. Kuwonetsetsa kuti mukutetezedwa tidzagwilitsa ntchito nambala ya chinsinsi imene idzayimile dzina lanu. Nambala imeneyi idzagwilitsidwa ntchito pa magazi komanso pa mikodzo yomwe mutazapeleke. Simukuyenera kupereka dzina lanu, keyala yanu ngakhalenso nambala yanu ya telefoni kwa ogwila ntchito za kafukufuku. Uthenga ulionse okhudza inu udzasungidwa m'malo otetezedwa bwino lomwe. Okhawo ogwila ntchito zakafukufuku ndi amene angazathe kuona uthenga okhudza inu . Dzina lanu komanso chilichonse chokhudza inu sichidzatuluka pa nthawi yotsindikiza zotsatira za kafukufukuyu.

Your records and samples will then be securely archived for future research studies. Future research may be done by the current research team or by other research teams working in other countries. Only your

records, samples and results which do not have your name or other identifiable information will be stored and shared with other research teams.

Zonse zokhudza inu zidzasungidwa ku malo otetezedwa kuti zidzathe kugwilitsidwanso ntchito mu kafukufuku winawa mtsogolo muno. Kafukufuku wa mtsogoloyo atha kudzapangidwa ndi ogwila ntchito za kafukufuku omwenso akupanga kafukufuku wapanoyu kapena ogwila ntchito za kafukufuku ochokera m'mayiko ena choncho uthenga omwe mupeleke pano udzatha kugwilitsidwa ntchito kupatulapo dzina lanu kapena chilichonse chokudziwitsani inu.

Your samples will be sent for testing in the central laboratory for this study which is located in Lilongwe and Karonga, and for testing in laboratories located in other countries outside of Africa. It is necessary to send some of your samples to specialised laboratories in other countries such as the UK and US because some of the tests we want to do cannot be easily carried out in Africa. The information linking your name to the unique code number will be locked securely with strict access control and will not be made available to the researchers working on your sample. Remaining portions of your samples will be stored for an unlimited period of time for future use in other studies of health and disease.

Magazi omwe mudzapeleke ngati mbali imodzi ya kafukufukuyu adzalembedwa nambala ya chinsinsi ndipo sipadzalembedwa dzina lanu. Magazi amenewa adzatumizidwa kuti akayezedwe ku laboratory amene ali ku Lilongwe ndi ku Karonga komanso ku ma laboratory ena akunja kwa Afilika. Ndi chinthu chofunika kutumiza ena mwa magazi anu ku mayiko ena ngati ku UK chifukwa zoyeza zina zomwe tikufuna tiyeze zitha kukhala zovutirapo kuti ziyezedwe kuno ku Afilika. Zonse zokudziwitsani inu komanso nambala yachinsinsi yomwe mudzapatsidwe zidzasungidwa mwachinsisi komanso motetezedwa ku malo osungika bwino ndipo ngakhale anthu opanga kafukufuku sadzakhala ndi mwayi owona zimenezi. Zikatha kuyezedwa magazi ena otsala adzasungidwa kwa kanthawi kuti adzathe kugwilitsidwa ntchito mu kafukufuku 378ioimpe mtsogolo muno wa zaumoyo.

12. Who can I contact if I have questions or complaints about the study?
Ndi ndani amene ndizamuyimbire ngati ndili ndi mafunso kapena madandaulo okhudza kafukufukuyu?

We would like to answer all your questions. If you have any questions now, please ask us. If you have any questions in the future, you can contact MEIRU, PO. Box 46, Chilumba (01364200/0999970392). If you have any further concerns, you can also contact the Malawi National Health Science Research Committee (01 789 400), who know of this study and can act on your behalf.

Tifuna kuyankha mafunso anu onse. Ngati muli ndi mafunso, chonde funsani. Ngati muzakhale ndi mafunso mtsogolo, mutha kuyimbila MEIRU, P.O. Box 46, Chilumba (01363200/0999970392). Ngati muli ndi zovuta zina muthanso kuyimba ku National Health Science Research Committee (01789 400), amene akudziwa za kafukufukuyu ndipo azakuthandizani.

Appendix 13. Information sheet English and Tumbuka (Bonje)

MEIRU
Malawi Epidemiology and Intervention Research Unit

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

MALAWI EPIDEMIOLOGY AND INTERVENTION RESEARCH UNIT POLOGALAMU YA LSHTM YA KAFUKUFUKU WA MATENDA OSAPATSIRANA

STUDY INFORMATION SHEET TUMBUKA

WANDUZGI WAWAKUTOLAPO LWANDE TUMBUKA

Study title: Chronic Kidney Disease of unknown cause (CKDu) in disadvantaged communities in low-and-middle income countries (LMICs)

Chonde werengani kapepala aka makora. Kakulongosola za kafukufuku uyo wakuchitika sono tifumbenge wowviri winu wakutolapo lwande. Chonde fumbani wakufukufuku kuti wamulongosolelani vya kafukufuku pala mukukhumba kumanya vinandi.Kutolapo lwande khwakujipereka ndipo kwali mwazomera kujira panji kutondeka kunjira ivi vipangiskenge yayi kuti muleke kupokera wowviri winu.

1. **Chifukwa wuli kafukufuku uyu wachitikenge?**

   Bakafukufuku waku Malawi Epidemiology na Intervention Research Unit (MEIRU) na wasukulu ya nkhwatha ya London School of Hygiene and Tropical Medicine wakuchita kafukufuku wa matenda gha figho agho ghakumanyikwa chala na umo ghakwambila, uyo waachitikenge kuno ku Malawi kweniso na vyaro vinyakhe vyapasi ndipo mukupempheka kutolapo lwande.

   Apa ninyengo iyo figho pa munthu yikuleka kugwira mwakwenerela, kweni vikumanyikwa yayi kuti vikutondeka wuli kugwira ntchito.Tikukhumba kumanya kuti matenda agha nganandi wuli, ntchifukwa wuli wanthu wakululwala , tikukhumbaso kumanya chomene suzgo ilo lilipo kwa wena Malawi.

2. **Kasi wovwiri wandalama wakupangira kafukufuku uyu wukufuma kochi kweneso kasi mbajani wakendeskero kakafukufuku awo wakazomerezga za kafukufuku uyu?**

   Kafukufuku watolenge vyaka viwiri ndipo wakupereka wowviri mba Medical Research Council, awo wakupereka wowviri wa wulere kukhala ku UK. Kafukufuku uyu wazomerezgeka na National

Health Sciences Research Committee muno mu Malawi na wa London School of Hygiene &Tropical Medicine Ethics Committees.

3. **Ntchifukwa wuli mwafumbika kutolapo lwande?**

Tikupempha wanthu wazaka kuposela 18 wakukhala ku Area 25 ku Lilongwe na wa ku Chilumba ku Karonga kutolapo lwande mukafukufuku. Werangani wanduzgi wakafukufuku ndipo mungafumba mafumbo agho muli nagho.

4. **Kasi kutolapo lwande mukafukufuku khwakujipereka?**

Lingawa khumbiro linu kutolapo lwande mukafukufuku. Mungalekezga nyengo yili yose. Wowviri winu wa chipatala uzamulekeka yayi chifukwa cha fundo yinu iyo mwakaka

5. **Kasi chikung'anamulachi kutolapo lwande?**

Tizakumfumbani mafumbo gha kukhwaskana na moyo winu (mwachiyerezgero kukhwewa hona na kamwero ka mowa, masewero yakukhwimiska thupi, na ulyero). Tizakumpimani uzito wa thupi, msinkhu, mu chiuno, na bio impedance iyo ya mkuona unandi wa mafuta na nyama thupi lino

Bakafukufuku wa zakumpephani kuti mulute ku chipatala cha ku area 25 na ku Chilumba kwambula kurya kalikose – ichi chikung'anamula kuti kuleka kurya na kumwa kalikose kufuma usiku.Muzamupepheka kupereka ndopa zichoko waka. Mungakana kupereka ndopa pala mwakhumba. Pala mwazomera tizamutola ndopa zichoko waka pa supuni (14 ml) Tizamuwika ndopa izo mutimabotolo tuchokotuchoko tu tatu.Wanthu wamanyi ntchito yakutola ndopa ndiwo wazamugwira ntchito iyi. Vilwero ivo vizamugwiriskika ntchito yakutola ndopa vizamkuwa vyakupwerereka makora ndipo pala wamala kugwiriska ntchito wazamuvitaya.Tu ndopa tuchoko wazamupimamo ka chibungu ka HIV. Tu ndopa tuchoko wazamupima umo figho yikugwirira ntchito.

Pala tamala ivi tizakumfumbani kuti mupereke matuzi pachoko(30ml – mu kabotolo ka Vaseline) kuti tizakapime umo figho zikugwirira ntchito.

6. **Chifukwa nivichi ndopa zitolekenge mukafukufuku uyu ndipo nivichi icho chichitikenge ku ndopa izi?**

Matenda ya Figho yakukhwaska umo Figho zinu zikugwirira ntchito. Tizamupima umo Figho zinu zikugwikira ntchito pakusefa ndopa.Tizamuwonaso matenda yanyakhe nga HIV kusazgiraso apa tizamusungako ndopa kuti tizakapimeko kuti tiwone ivyo vikwambiska matenda yambula kupasirana.

11. Kasi malonje ghakukhwaskana na ndopa na matuzi agho ine nipelekenge ghasungikenge mwa chisisi uli?

Malonje ghose agho tipulikenge kufuma kwa imwe ghakukhwaskana na kafukufuku uyu ghasungikenge mwa chisisi chomeni. Pakuvikilira chisisi chinu titigwiliskenge ntchito nambara gha chisisi kwimilira imwe na malonje ghinu pamo na ndopa na matuzi agho titolenge. Mwanthowa yiliyose zina, uko mukufuma panji nambara yinu ya phone yipelekekenge kulikose yayi. Tisungenge ivi mwa kuvijalira na loku. Pachifukwa cha kafukufuku uyu pera, gulu la wakufufuza la kafukufuku uyu wawonenge malonje na ivyo tatola. Zina linu na fundo zinyakhe ivyo vingalongora kuti ndimwe vizamusangikamo yayi panyengo ya kuthandazga vyakulondezgapo vyakhe.

Pachifukwa cha wakafukufuku wa kunthazi fundo na ivyo tatola visungikenge mwa pamoza mwakupwelelereka.Bakufukufuku bakunthazi bangazakachitika na gulu leneliri panji na gulu linyakheso mu vyaro vinyakhe. Ni fundo zekha izo zilije zina linu nesi malonje agho ghangavumbura kuti ndimwe agho ghasungikengi na kugawanako na magulu gha kafukufuku ghanyakhe.

Ivyo titolenge kwa imwe titumizgenge kuti wakapimekoso ku Lab yikuru kungawa gha ku Lilongwe panji ku Karonga nanga ni ma lab ghaku vyaro vyakuwaro kwa Africa. Ntchakukhumbikwira kutumizgako ivi ku ma Lab ghapaderadera ku vigawa nga ndi Uk na USA chifukwa vinji ivyo tikukhumba kupimako vingapimika yayi mu vyaro vya mu Africa. Fundo zakuvumbula zina linu na nambara yinu yachisisi vibisikenge kwa agho wakugwirapo pa ndopa na matuzi ghinu. Ivyo vikhalengepo visungikenge kwa nyengo yambura kumara pa chifukwa cha ntchito za kunthazi za umoyo na matenda.

12. Usange pali mafumbo panji madandaulo za kukwaskana na kafukufuku uyu ningaluta kwa njani?

Tingatemwa kuzgora mafumbo ghinu ghose. Usange pali mafumbo sono chonde tifumbani. Usange munthazi mwazakawa na mafumbo mungakakhwaskana na MEIRU, PO Box 46 Chilumba. (01364200/0999970392). Usange mungawa na madandaulo munthazi mungazakawonanaso na wa Malawi National Health Science Research Committee (+265 888 344 443) iwo wakumanya za kafukufuku uyu ndipo wangamilira mumaro mwa imwe.

Ndopa izo muzamupereka zizamukugwiriskika ntchito mukafukufuku uyu pera. Zizamuyeghekera ku laboratory kuti tikazipime .Vyakulondezga vyose vizamusungika mwachisisi

7. **Kasi mungavimanya ivyo tikupima kweniso kuwona vyakulondezga vya ndopa zinu?**

Tingamuphalirani pa ivyo tamupimani (mwachiyerezgero uzito winu),kuchimbira kwa ndopa zinu panyengo yeneyiyo.Pala mukukhumba tingakhwaskana namwe pala pali vyambula kupulikiskika pavyakulondezga vya ndopa zinu. Tingazakayimba foni,Bakugwira ntchito mukafukufuku uyu bazamudumba namwe icho chikung'anamula pavyakulondezga vinu.Muzamuphalilika pala ntchakwenerera kuti mukapimikeso kuchipatala. Pala muzamutumizgika kuchipatala chikulu panji chichoko,bakugwira nthito kuchipatala bazamudumba namwe kumulangizani pakusintha makhaliro yinu panyakhe kupereka makhwala kuti umoyo winu uwe makola.Pala vyakulondezga vya ndopa vili makola , tikhwaskana namwe yayi, kweni nipera mungatifumba kuti musimikizge.

8. **Kasi ningakana kuba mukafukufuku uyu panyakhe kukana kupereka ndopa?**

Niwanangwa winu kusankha kunjira mukafukufuku uyu panji kuleka kunjira. Mungasankha kuleka kuba mukafukufuku uyu panji kukana kupereka ndopa panyengo yili yose kwambula kupereka chifukwa. Ichi chizamukhuzana yayi nawovwili uwo mungapokera kuchipatala.

9. **Kasi uweni wakutolapo lwande mukafukufuku uyu ni vichi?**

Usange vyakulondezgapo vinu vili makora viwi yayi titimutumizganinge ku chipatala kuti mukapimiskekoso vipimo vinyakhe uko mungakapokera maluso ya makhaliro gha umoyo uwemi panji mankhwara gha kovwira kuti mukhalenge makola.

Kutolapo lwande kwinu kuwovwirenge wakufufuza kuwa na kapulikiskiro kawemi ka matenda gha fiwu agho kwamba kwake kundamanyikwe. Kutolapo lwande kwinu kuwovwirenge kupulikiska vinji vyakwambiska nthenda iyi kweniso ntchifukwa uli pali kupambana pakati wanthu awo wakukhala mu mizi na awo wakukhala mu tauni. Ichi chiwovwirenge mwakuluska kajivikililiro na kapwelelero ka nthenda iyi.

10. **Kasi ulwani wakunjira mu kafukufuku uyu ni vichi?**

Pakutolapo lwande mu kafukufuku uyu mukumanenge na ulwani uli wose yayi. Pakutola ndopa mumusempha mungapulikako ulwilwi pachoko, kweni ndi ulwilwi wakulutilira yayi.pamalo apo nyereti yangugwaza pangapanga ka bamba. Palije icho imwe mulipilengepo chifukwa chakuti mwatolapo lawande.

Appendix 14: Informed consent documents

English

MALAWI EPIDEMIOLOGY AND INTERVENTION RESEARCH UNIT (MEIRU)

CHRONIC KIDNEY DISEASE OF UNKNOWN ORIGIN CONSENT FORM – ENGLISH

Your Consent:
Please answer the questions below to show that you agree to take part in the study. After reading the statements, you will be asked to sign using a pen or your thumb print. You will also be required to write down your name together with todays date at the bottom of the page. The interviewer will also counter sign. A copy of this consent will be kept by the interviewer and another will be given to you to keep.

Study ID	*Study ID*	STID

Date completed	*Date*	INTDATE

1	I have read or have been read the information concerning this study and I have been given the study information sheet. I have been given the chance to ask all the questions to help me understand why this study is being done. I know that my participation is voluntary, that I can stop being in the study at any time without giving a reason and that the information collected will be kept strictly confidential. I know that the final results will nott contain any names.	1=Yes CONSENTA 2= No If CONSENTA=2, Skip to RCDR
2	I agree to take part in this study and to answer questions about my health and lifestyle. I also agree to have my weight, height, waist and hip size, blood pressure and pulse.	1=Yes CONSENTB 2=No If CONSENTB=2, Skip to RCDR
3	I agree to give a sample of my blood (<10ml) to test for my renal function. I know that I have been asked to not eat anything for four hours before giving my blood. I also agree to give a small amount of urine (30 mls) that will also be used to see how my kidneys are functioning.	1=Yes CONSENTC 2= No
4	I agree to be tested for HIV. If I am found to be HIV positive, I will be ask to repeat the test at the nearest health centre.	1=Yes CONSENTD 2=No

		3=Not applicable
4	I agree to be tested for HIV. If I am found to be HIV positive, I will be ask to repeat the test at the nearest health centre.	1=Yes CONSENTD 2=No 3=Not applicable
5	I agree that a minimal amount of my blood (<2mls) can be stored for further research on this disease.	1=Yes CONSENTF 2=No Skip if CONSENTC=2

Repoting of blood results – Please sign or thumb print to indicate that you would like to be told the results of the blood test if they are abnormal. If the results are normal we will not contact you but you can still contact us to confirm.

I would like to be told the results of my blood tests if they show signs of kidney disease.	1= Yes DISCLOSURE 2= No **Skip if CONSENTC=2**

Name of participant		
First name (s)		
Surname		

	Interviewer code		RCDR

Signature of thumb print Name Date

Eyewitness Name Date

Interviewer Name Date

Appendix 15. Informed consent document (Chichewa, for Area 25 survey)

MALAWI EPIDEMIOLOGY AND INTERVENTION RESEARCH UNIT (MEIRU)

CHRONIC KIDNEY DISEASE OF UNKNOWN ORIGIN CONSENT FORM – CHICHEWA

Chilolezo Chanu:

Ngati musatenge nawo gawo mu kafukufuku, tikupemphani kuti muvomereze pakuyankha mafunso ali mmunsiwa. Mufunsidwa kusaina kapena kudinda ndi chala chanu cha mkomba phala. Mufunikans kulemba dzina lanu , sayini yanu(kapena chidindo cha chala chanu cha mkomba phala) ndi tsiku la lero pansi pa tsambali. Ifenso tidzasayina ndikulemba tsiku la lero pa tsambali. Ife tisunga pepalali komanso tikupatsani pepala lina lofanna ndi ili kuti musunge.

Study ID	*Study ID*	STID

Date completed	*Date*	INTDATE

1	Ndawereng kapena andiwerengera zofunik pa kafukufuku ndipo ndapatsidwa pepala la uthenga wa kafukufukuyu kuti ndisunge.Ndapatsidwa mwayi ofunsa mafunso kuti ndidziwe cholinga cha kafukufukuyu.Ndikudziwa kuti siwokakamiza ndipo nditha kusiya nthawi ina ili iliyonse ndipo chithandizo changa zotsatira zakafukufuku zizasungidwa mwachinsinsi komanso motetezedwa bwino lomwe ku Malawi. Ndikudziwa kuti zotsatira zake zidzasindikizidwa osatchula mayina.	1=Eya CONSENTA 2= Ayi If CONSENTA=2, Skip to RCDR
2	Ndavomera kutenga nawo gawo pa kafukufuku ndikuyankha mafunso okhuza umoyo ndi khalidwe langa.Ndavomera kutenga utali wanga, kulemera, kutalika kwa ntchafu ndi chiwuni, kuthamanga kwa magazi ndi kugunda kwa magazi mu kafukufuku.	1=Eya CONSENTB 2=Ayi If CONSENTB=2, Skip to RCDR
3	Ndavomera kupereka magazi okwana supuni yaikulu imodzi kuti ayezedwa kugwila ntchito kwa ipsyo. Ndikudziwa kuti ndafunsidwa kusala chakudya kwa maola asanu ndi atatu asanatenge magazi.	1=Eya CONSENTC 2=Ayi
4	Ndavomera kuyezetsa matenda a HIV. Ngati ndapezeka ndi HIV, ndidzafunsidwa kupita kuchipatala cha m'dera lathu kuti ndikayezedwenso.	1=Eya CONSENTD 2=Ayi 3=Not applicable

5	Ndavomera kuti magazi anga ochepa (<2ml) atengedwe ndikusungidwa (nthawi yosadutsa zaka zisanu) kuti adzagwiritse ntchit kuyeza matenda ena ofunika m'malawi	1=Eya CONSENTF 2=Ayi Skip if CONSENTC=2

Kuulula zotsatira za magazi – Chonde sayinani kapena dindani nd chala chamkomba phala kusonyeza kuti mukufuna kudziwa kuti mumve zotsatira za magazi anu ngati mwapezeka ndi matenda.Ngati zotsatira zanu zili bwino sitizakupezani koma mutha kutsata nokha kudzatenga zotsatirazo.

Ndikufuna kuuzidwa zotsatira za kugwila ntchito kwa ipsyo zanga ngati zisonyeza matendawa	1= Eya DISCLOSURE 2= Ayi **Skip if CONSENTC=2**

	Dzina la wotenda nawo gawo		
	First name (s)		
	Surname		

	Interviewer code		RCDR

Sayini kapena chidindo cha chala cha mkomba phala Dzina Tsiku

Wochitira umboni Dzina
Tsiku

Ofunsa mafunso Dzina
Tsiku

Appendix 16. Informed consent document, Tumbuka (Bonje)

MEIRU
Malawi Epidemiology and Intervention Research Unit

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

MALAWI EPIDEMIOLOGY AND INTERVENTION RESEARCH UNIT (MEIRU)

CHRONIC KIDNEY DISEASE OF UNKNOWN ORIGIN CONSENT FORM – TUMBUKA

Kuzomelezga Kwinu:
Chonde zgolani mafumbo agha kuti mukhozgele kutolapo lwande kwinu mu kafukufuku. Usange mwaŵerenga fundo izi, mufumbikenge kusayinila na chakulembela panji chidindo cha chigunyu. Mukhumbikwengeso kulemba zina linu pamoza na deti la muhanyauno pasi pa pepala ili. Wakufumba nayo wasayinilenge. Kalata yimoza yakuzomelezga kwinu yisungikenge na uyo wamufumbaninge yinyakhe mupasikenge imwe kuti musunge.

Study ID	*Study ID*	STID

Date completed	*Date*	INTDATE

1	Naŵerenga panji ŵaniŵerengela wanduzi wakukhwaskana na kafukufuku uyu ndipo ŵanipasa kalata ya wanduzi. Napasika mwaŵi wakufumba mafumbo ghose kuti nipulikiske chifukwa icho kafukufuku uyu wakuchitikila. Nkhumanya kuti kutolapo lwande kwane nkhwakujipeleka, kuti ningaleka kuŵamo mu kafukufuku nyengo yiliyose kwambula kupeleka chifukwa ndipo wanduzi uwo wukatoleka wamusungika mwachisisi nkhanila. Nkhupulikiska kuti vyakumanyikwila vyane ngati zina, nambala ya foni na nyumba vyamuvumbukwa chala ku ŵanthu ŵakuwalo kwa kafukufuku uyu. Nkhumanya kuti vyakusangika vyaumaliro vamkuŵavya mazina ghalighose.	1= Enya CONSENTA 0= Yayi **If CONSENTA=0, Skip to RCDR**
2	Nkhuzomela kutolapo lwande mu kafukufuku uyu na kuzgola mafumbo ghakukhwaskana na umoyo wane na ukhaliro wane. Nkhuzomelaso kuti uzitu wane, utali, chiuno, kuchimbila kwa ndopa na bio impedance.	1=Enya CONSENTB 0=Yayi **If CONSENTB=0, Skip to RCDR**
3	Nkhuzomela kupeleka ndopa (zakuchepela 14ml) kuti nipimike umo figho zikugwilira ntchito. Nkhumanya kuti napempheka kuleka kurya kalikose kwa maora 8 pambele nindapeleke ndopa. Nkhuzomelaso kupeleka matuzi (30 mls) agho ghamugwiliskika ntchito kukawona umo figho zane zikugwilira ntchito.	1= Enya CONSENTC 0= Yayi

4	Nkhuzomela kuti nipimike HIV. Usange nasangika kuti nilinayo HIV, nifumbikenge kuti nkhapimiskeso ku chipatala cha pafupi chomene.	1= Enya CONSENTD 0=Yayi 3= Fumbo lambula kwenelera
5	Nkhuzomela kuti ndopa zane zichoko (zakuchepela 2mls) zingasungika pa chifukwa cha kulutilizga kafukufuku pa nthenda iyi.	1= Enya CONSENT E 0=Yayi **Skip if CONSENTC=0**
6	Nkhupulikiska kuti wanduzi uwo napeleka wungakasungika na awo ŵakupanga kafukufuku, nkhupulikiska kuti wanduzi uwo napeleka ungagwiliskika ntchito mu vyakuthandazga, malipoti, pa neti, na vyakufumiska vinyakhe vya kafukufuku mu kaŵilo kambula ine kumanyikwa. Nkhupulikiska kuti awo ŵakupanga ŵakafukufuku ŵanyakhe ŵangaŵa na mwaŵi wakuwona wanduzi uwu mu kaŵilo ka kuvikilirika kambula ine kumanyikwa.	1=Enya CONSENT F 0=Yayi

Kuphala vyakulondezgapo vya ndopa – Chonde sayinani panji mudinde na chigunyu kulongola kuti mukukhumba kuzakaphalirika vyakulondezgapo vya kupimika ndopa usange pali kaheni.

Nkhukhumba kuzakaphalirika vyakulondezgapo vya kupimika kwa ndopa zane usange zikulongola vimanyikwilo vya nthenda za figho.	1= Enya DISCLOSURE 0= Yayi **Skip if CONSENTC=0**

Zina la wakutolapo lwande		
Zina		
Chiwongo		

	Interviewer code		RCDR

Sayini panji chidindo cha chigunyu Zina Dazi

Wakufumba Zina Dazi

Kaboni Zina Dazi

Appendix 17: DEGREE Core questionnaire

DEGREE Core Protocol Porforma

Consent, Interview Language and Name	Response	Code
Consent has been read and obtained	Yes 1 No 2 If NO, END	I5
Interview Language *[Insert Language]*	English 1 *[Add others]* 2 *[Add others]* 3 *[Add others]* 4	I6
Time of interview (24 hour clock)	⌊_⌋⌊_⌋ : ⌊_⌋⌊_⌋ hrs mins	I7
Family Surname		I8
First Name		I9
Additional Information that may be helpful		
Contact phone number where possible		I10

Question	Response	Code
Sex (*Record Male / Female as observed)*	Male 1 Female 2	C1
What is your date of birth? *Don't Know 77 77 7777*	d' mm year	C2
How old are you?	Years	C3
In total, how many years have you spent at school and in full-time study (excluding pre-school)?	Years	C4
What is your *[insert relevant ethnic group / racial group / cultural subgroup / others]* background?	[*Locally defined]* 1 [*Locally defined]* 2 [*Locally defined]* 3 Refused 88	C6
Which of the following best describes your main work status over the past 12 months?	Government employee 1 Non-government employee 2 Self-employed 3 Non-paid 4 Student 5 Homemaker 6 Retired 7 Unemployed (able to work) 8 Unemployed (unable to work) 9 Unpaid domestic 10 Refused 88	C8
If you are working what is your main occupation *[FREE TEXT]:*		OCCTXT
Do you have experience of migrant work? [Defined as staying far from home for seasonal work]	Yes 1 No 2	MIGR
Can you give an estimate of the monthly household income if I read some options to you? Is it	≤ Quintile (Q) 1 1 More than Q 1, ≤ Q 2 2	C11

[INSERT QUINTILE VALUES IN LOCAL CURRENCY] *(READ OPTIONS)*	More than Q 2, ≤ Q 3 More than Q 3, ≤ Q 4 More than Q 4 Don't Know Refuse'	3 4 5 77 88	
During the past 12 months, how frequently have you had at least one standard alcoholic drink? *(READ RESPONSES, USE SHOWCARD)*	Daily 5-6 days per week 3-4 days per week 1-2 days per week 1-3 days per month Less than once a month Not at all Refused	1 2 3 4 5 6 7 88	A4
Do you currently smoke any tobacco products, such as cigarettes, cigars or pipes? *(USE SHOWCARD)*	Yes No	1 2	*T1*
In a typical week, on how many days do you eat MEAT *(USE SHOWCARD)*	Number of days Don't Know 77	⌊_⌋⌊_⌋ ,	*D1*
Does your work involve vigorous-intensity activity that causes large increases in breathing or heart rate like *[carrying or lifting heavy loads, digging or construction work]* for at least 10 minutes continuously? *(OR USE SHOWCARD)*	Yes No	1 2	P1
Have you ever been told by a doctor or other health worker that you have raised blood pressure or hypertension?	Yes No	1 2	H2a
Have you ever been told by a doctor or other health worker that you have raised blood sugar or diabetes?	Yes No	1 2	H7a

DEGREE study core physical measurements

Question	Response		Code
Ambient Temperature (at time of examination		⌊_⌋⌊_⌋⌊_⌋°C	Temp
Blood Pressure			
Interviewer ID		⌊_⌋⌊_⌋⌊_⌋	M1
Device lood pressure		⌊_⌋⌊_⌋⌊_⌋	M2
Cuff size us	1 Medium Large	M3 2 3	
Reading 1	Systolic (mmHg)	⌊_⌋⌊_⌋⌊_⌋	M4a
	Diastolic (mmH⌊_⌋	M4b	
Reading 2	Systolic	⌊_⌋⌊_⌋⌊_⌋	M5a
	Diastolic (mmHg)	⌊_⌋⌊_⌋⌊_⌋Reading 3 Systolic (mmHg)	
	Diastolic (mmHg)	⌊_⌋⌊_⌋⌊_⌋	M6b
he past two weeks, have youeated for raised blood pressure with drugs	Yes No	1 2	M7
Height, Weight and BIA			

For women: Are you pregnant?	Yes	1	M8a
	No	2	
For women: Are you mensturating?	Yes	1	M8b
	No	2	
Have you eaten 391ioimpedancees	1	M9	
	No	2	
Interviewer ID		⌊_⌋⌊_⌋⌊_⌋	M10
Height	in Centimetres (cm)	⌊_⌋⌊_⌋⌊_⌋ M11	
Device ID for BIA		⌊_⌋⌊_⌋⌊_⌋	M13
BIA Device makeel (50Hz supine)			M14
Bioi value		⌊_⌋⌊_⌋⌊_⌋ Ω	BIA1
Bioimpedance outputs in kg or % *(tick)*		☐kg ☐%	BIA2
Bioimpedance outputs	Fat Mass	⌊_⌋⌊_⌋⌊_⌋	BIA3a
different machines may provide different outputs	OR FMass (also termed lean mass)	⌊_⌋⌊_⌋⌊_⌋	BIA3 b

Appendix 18: DEGREE SOP for specimen collection in Malawi sites

SPECIMEN COLLECTION

URINE

A fresh, cleanly collected midstream 10 -20ml urine sample is required.

The containers for sample collection should be clean, dry, leak-proof and sufficiently wide neck for easy use by the patient. It must be free of all traces of disinfectants, elements, however, need not be sterile. Transport urine specimen in biohazard bag in cooler box. Can be stored at 2 – 8oC if analysis is to be delayed. But it is advisable to analyze the specimen as soon as possible for tests like glucose.

BLOOD

Blood samples are collected by KPS/MEIRU field staff. The blood samples should be collected using the evacuate method; using a vacutainer tube. This method consists of a plastic hub, a hypodermic needle and a vacuum tube. This will allow blood to flow freely into a vacutainer tube.

The following tubes will be used to collect blood: -

EDTA vacutainer tubes (purple top), Fluoride Oxalate Tube (grey top), Plain Tubes (red top)

Three tubes will be collected per participant. The volume of samples per client in total will be 18mls; 6 ml EDTA, 2mls Fluoride oxalate and 10 ml clotted red top sample.

Specimen tubes should be compatible with automated systems and instruments used by the laboratory.

EDTA tube

Aim- the EDTA sample will keep whole blood unclotted for Full Blood count. It will preserve the cells in blood intact so that they could be well counted. Specimens to be processed for hematology must be run within 24 hours after collection if they have been kept at room temperature and 36 hours if kept at 4o centigrade. Make sure blood is properly mixed by gently inverting the tube end-over-end 8 to 10 times, soon after collection and before transporting. Keep and transport EDTA samples in a cooler box, unrefrigerated. Deliver the specimen to the lab as soon as possible to be tested within 6 hours from time of collection (i.e. same day) EDTA tubes should at least be over half filled with samples.

Flouride Oxalate (grey top)

Aim- The aim of Fluoride oxalate tube is to keep whole blood unclotted and preserve it to avoid glycolysis. Fluoride inhibits enolase, the enzyme involved in glycolysis.

Make sure blood is properly mixed by gently inverting the tube end-over-end 8 to 10 times.

Keep and transport Fluoride Oxalate samples for glucose in a cooler box, refrigerated i.e. ice packs if delay is **anticipated**. Deliver the specimen to the lab as soon as possible to be tested within 6 hours from time of collection (i.e. same day)

Serum (red top)

Aim- Serum samples will be collected for chemistry analysis and storage.

Blood will be collected into the red top tubes (without anticoagulant) to obtain serum. Sample collection for serum specimens is done as per routine phlebotomy procedure.

Venepuncture Procedure

Identify the patient: ask the patient to state his/her name and his/her birth date. Check these against the requisition form (paper or electronic).

Check the requisition form for requested tests, other patient information and any special draw requirements.

Gather the tubes and supplies that you will need for the draw.

Position the patient in a chair or sitting or lying on a bed.

Put on non-latex gloves, fix the needle to the vacutainer holder (Plastic hub)

Select a suitable site for venepuncture, by placing the tourniquet 3 to 4 inches above the selected puncture site on the patient.

Ask the patient to make a fist; avoid “pumping the fist”. Grasp the patient’s arm firmly using your thumb to draw the skin taut and anchor the vein. Swiftly insert the needle through the skin into the lumen of the vein. The needle should form a 15-30-degree angle with the arm surface. Avoid excess probing. The bevel of the needle should face away from you

Hold tube in vacutainer holder and draw blood into red top tube before filling into other tubes with anticoagulant, one after the other. This will prevent anticoagulant contamination.

When the last tube is filling, remove the tourniquet.

Remove the needle from the patient's arm using a swift backward motion.

Place gauze/cotton immediately on the puncture site. Apply an’ hold adequate pressure to avoid formation of a hematoma. After holding pressure for 1-2 minutes, tape a fresh piece of gauze or Band-Aid to the puncture site.

Order of Draw of Blood:
Blood collection tubes must be drawn in a specific order to avoid cross-contamination of additives between tubes. The recommended order of draw for plastic vacutainer tubes is:
non-additive tube (red top)
Fluoride oxalate tube (Gray top)
EDTA (Lavender or purple top) tube.

Labeling the Sample
Stick one SPECNO label onto the specimen form and one onto the sample tube. In the sample tubes, try to copy participants initials, sex, DOB and STID if possible

Fingerstick Procedure:
Follow steps #1 through #4 of the procedure for venepuncture as outlined above.
The best locations for fingersticks are the 3rd (middle) and 4th (ring) fingers of the non-dominant hand. Do not use the tip of the finger or the center of the finger. Avoid the side of the finger where there is less soft tissue, where vessels and nerves are located, and where the bone is closer to the surface. The 2nd (index) finger tends to have thicker, callused skin. The fifth finger tends to have less soft tissue overlying the bone. Avoid puncturing a finger that is cold or cyanotic, swollen, scarred, or covered with a rash.
1. When a site is selected, put on gloves, and cleanse the selected puncture area.
2. Massage the finger toward the selected site prior to the puncture.
3. Using a sterile safety lancet, make a skin puncture just off the center of the finger pad. The puncture should be made perpendicular to the ridges of the fingerprint so that the drop of blood does not run down the ridges.
4. Wipe away the first drop of blood, which tends to contain excess tissue fluid.
5. Collect drops of blood into the collection tube (capillary)/device by gentle pressure on the finger. Avoid excessive pressure or “milking” that may squeeze tissue fluid into the drop of blood.
6. Rotate and invert the collection device to mix the blood collected (optional).
7. Have the patient hold a small gauze pad over the puncture site for a few minutes to stop the bleeding.
8. Dispose of contaminated materials/supplies in designated containers.

Appendix 19: Transportation of samples and laboratory analysis

Chemical analyses of bio samples were conducted in regional MEIRU-run laboratories. Bonje samples were sent to the nearby Karonga Prevention Study (KPS) laboratory, and Area 25 samples were sent to the MEIRU headquarters laboratory in Lilongwe.

Urine samples

After collection of urine samples, the labelled test tubes were immediately placed in a cold box with dry ice packs (4-8^{o}c) and transported by field staff to the respective laboratories for each study zone for centrifugation and storage within 6 hours of collection. The laboratory technician receiving the urine samples cross-checked the number of tubes per participant, sample volume, specimen forms, rapid test forms and ID numbers. If the laboratory technologist found specimens that did not adhere to the DEGREE standard operating procedures (SOP) (appendix E), this was noted on a log form, the sample was discarded, and the team leader was informed to return and retake the sample. Urine samples were analysed immediately after arrival at the laboratory, and two aliquots of 2ml samples were transferred to 2mk RNase free cryotubes and stored in a -80^{o}C lab freezer. Each stored aliquot was labelled, and data entered into the laboratory database for potential future studies.

Blood samples

After blood samples were collected, the labelled test tubes were placed upright for 10 minutes to allow for clotting. They were then packed in a cool box (as described for urine samples) and transported to the regional laboratory for centrifugation and storage within 2 hours of collection. The laboratory technician receiving the blood samples cross-checked the number of tubes per participant, sample volume, specimen forms rapid test forms and ID numbers. If the laboratory technologist found specimens that did not adhere to the SOP (appendix E), this was noted on a log form, the sample was discarded, and the team leader was informed to return and retake the sample. Chemical analyses of bio samples were conducted in regional MEIRU-run laboratories. Bonje samples were sent to the nearby Karonga Prevention Study (KPS) laboratory, and Area 25 samples were sent to the MEIRU headquarters laboratory in Lilongwe. For transportation detail please see appendix F.

Bio-sample laboratory analyses

First, a dipstick urinalysis was performed on the urine samples using commercial testing sticks to report semi-quantitative values for blood, protein (at the macroalbuminuric level), leucocytes, glucose, pH and specific gravity in the sample.

Serum creatinine

Serum creatinine was measured in order to calculate the participant eGFR. Plain serum samples were centrifuged for ten minutes at 1300-2000 RFC (g) in a swing bucket centrifuge. Serum creatinine was measured in the separated serum using the compensated Kinetic Jaffé method calibrated to isotope dilution mass spectrometry (IDMS) standards (see appendix G for detail). Two 2ml aliquots of serum were transferred into an RNase cryotube and stored in a -80^{o}C freezer and logged for any future analyses. Urine creatinine was also measured using the Jaffé method.

Urine albumin

Urine albumin was measured using the photometric colour test method. The urine creatinine and urine albumin measures are used to calculate the albumin: creatinine ratio (ACR) which signifies whether the participant has proteinuria (protein in the urine) which is a clinical signifier of kidney disease.

Glucose and glycated haemoglobin

Blood glucose levels were measured to test for diabetes. The Fluoride/Oxalate blood samples were centrifuged using the same procedure as described above for plain serum. A 2ml aliquot of plasma was collected and a random glucose test was conducted using the hexokinase method, which is an enzymatic UV test (see appendix H for detail). The EDTA full blood samples were used to measure glycated haemoglobin (HbA1c %).

Appendix 20: Jaffe Method

In the compensated kinetic Jaffe (traceable to the IDMS reference) method, creatinine reacts with picrate ion formed in alkaline medium to develop a red-orange colour. The colour produced from the sample is then compared in a colorimeter at wavelength of 505nm with that produced by a known amount of creatinine under the same condition.
Creatinine analysis was conducted using an AU480 analyser with ISE (Beckman Coulter, California, USA), a fully automated, random access chemistry system equipped with STAT capability. The AU480 analyser automatically computed the creatinine concentration of each sample and interference from protein was mathematically auto corrected by the analyser by subtracting 18 µmol/L from each test result.

Appendix 21: Hexokinase method

The enzyme hexokinase (HK) catalyses the reaction between glucose and adenosine triphosphate (ATP) to form glucose-6-phosphate (G-6-P) and adenosine diphosphate (ADP). In the presence of nicotinamide adenine dinucleotide (NAD), G-6-P is oxidized by the enzyme glucose-6-phosphate dehydrogenase (G-6-PD) to 6-phosphogluconate and reduced nicotinamide adenine dinucleotide (NADH). The increase in NADH concentration is directly proportional to the glucose concentration and can be measured spectrophotometrically at 340nm. [https://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/l10am_c_met_glucose.pdf]

Appendix 22: Sample chemistry and reagents used

Sample type	Measure	Method used
Urine	Electrolytes	Crown ether membrane electrode
	Albumin	Turbidimetric method
	Creatinine	Compensated kinetic Jaffé method
	Urine albumin (UALB)	Turbidimetric method.
Serum	Total cholesterol	Enzymatic colour test
	HL cholesterol	
	LDL cholesterol	
	Triglycerides	Enzymatic colour test
	ASO	Immuno-turbidimetric test
	CRP	Immuno-turbidimetric test
	Creatinine	Compensated kinetic Jaffé method
	Urea	Kinetic UV test
	Uric acid	Enzymatic colour test
	AST	Kinetic UV test
	ALT	
	GGT	
	Electrolytes	Crown ether membrane electrode
EDTA whole blood	HbA1c %	HbA1c/Total Hemoglobin ratio expressed as percentage
	Total Hb (THb)	Immuno-inhibition test
Plasma	Glucose	Enzymatic UV test: hexokinase method

Appendix 23. Supplementary table showing original categories of socioeconomic and anthropometric measures, Malawi

Table 1. Sociodemographic and anthropometric characteristics of overall study participants (prior to exclusion of population with diabetes, hypertension and heavy proteinuria) (n=1076)

Variable	eGFR n= 1076		eGFR categories, n (%)[b] n= 5	n= 79	n= 992
	n (%)[a]	Mean (SD)	<60	>=60, <90	>=90
Age (years)					
<20	66 (6)	133.2 (14.6)	0	1 (2)	65 (99)
20-29	362 (34)	125.3(12.7)	0	3 (1)	359 (99)
30-39	280 (26)	115.4 (12.4)	1 (0.4)	5 (2)	274 (98)
40-49	185 (17)	106.9 (12.6)	3 (2)	5 (3)	177 (96)
50-59	79 (7)	100.7 (8.9)	0	12 (15)	67 (85)
60+	104 (10)	89.0 (11.9)	1 (1)	53 (51)	50 (48)
Sex					
Female	656 (61)	114.9(16.9)	4 (1)	40 (6)	610 (93)
Male	422 (39)	114.5 (18.0)	1 (0.2)	39 (9)	382 (91)
Education (years)					
≤5	108 (10)	101.9(18.4)	1 (1)	24 (22)	83 (77)
>5≤10	399 (37)	114.6(16.8)	2 (0.2)	27 (7)	370 (93)
>10	571 (53)	117.3(16.4)	2 (0.4)	28 (5)	539 (94)
Occupation					
Office worker	245 (23)	120.9(16.7)	1 (0.4)	8 (3)	236 (96)
Self-employed	133 (12)	114.1(16.2)	1 (1)	8 (6)	124 (93)
Farmer	493 (46)	111.5 (16.7)	3 (1)	45 (9)	445 (90)
Fisherman	18 (2)	104.8(13.2)	0	2 (11)	16 (89)
Unemployed	59 (6)	117.7(16.4)	0	3 (5)	56 (95)
Unpaid/domestic worker	113 (11)	119.4(17.5)	0	4 (4)	109 (95)
Retired	12 (1)	86.5 (11.7)	0	8 (67)	4 (33)
Refused	3 (0.3)	105.0(21.5)	0	1 (33)	2 (67)
Household monthly income (MK) [c]					
<5000	145 (14)	117.4(21.0)	0	14 (10)	131 (90)
>5,000 <10,000	177 (16)	114.9 (18.0)	1 (1)	15 (9)	161 (91)
>10,000 <20,000	193 (18)	114.8(16.8)	1 (1)	13 (7)	179 (93)
>20,000 <40,000	219 (20)	115.4(16.0)	1 (1)	12 (6)	206 (94)
>40,000	321 (30)	112.9(15.7)	1 (0.3)	23 (7)	297 (93)
Don't know	17 (2)	113.5 (24.2)	1 (6)	1 (6)	15 (88)
Refused	4 (0.4)	115.9 (24.9)	0	1 (25)	3 (75)
Area					
Urban (Area 25)	326 (30)	115.6(15.5)	1 (0.3)	15 (5)	310 (95)
Rural (Bonje)	750 (70)	114.4(18.1)	4 (1)	64 (9)	682 (91)
Healthy lifestyle choices					
Non-smoker/ alcohol drinker	820 (77)	114.6 (17.4)	5 (1)	60 (7)	755 (92)
Smoker/ alcohol drinker	256 (24)	115.2 (17.2)	0	19 (7)	237 (93)
Regular meat-eater					

Yes	247 (23)	113.5 (16.4)	0	21 (9)	226 (92)
No	829 (77)	115.1(17.6)	5 (1)	58 (7)	766 (92)
Body mass index (kg/m^2)					
Underweight (≤18.5)	55 (5)	119.3 (20.2)	0	6 (11)	49 (89)
Normal (>18.5 - ≤ 25)	677 (63)	116.7(16.9)	1 (0.1)	40 (6)	636 (94)
Overweight (>25 - ≤30)	242 (23)	111.0(15.8)	3 (1)	16 (7)	223 (92)
Obese (>30)	102 (10)	108.3(18.5)	1 (1)	17 (17)	84 (82)
Fat-free mass (kg)					
1st tertile (≤37)	171 (16)	110.2 (18.1)	0	26 (15)	145 (85)
2nd tertile (>37 - <45)	442 (41)	115.0(18.0)	4 (1)	27 (6)	411 (93)
3rd tertile (≥45)	463 (43)	116.2 (16.1)	1 (0.2)	26 (6)	436 (94)

[a] Percentages in columns; [b] percentages in rows; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol; [c] Exchange rate (MK to USD) 0.001 at time of questionnaire

Appendix 24. Supplementary table showing linear and logistic regression models for Bonje

Table 1. Linear and logistic regression models, showing both minimally and fully adjusted models Bonje (n=578)

Variable	Bonje Model 1 eGFR	Bonje Model 2 eGFR	Bonje Model 3 eGFR <90	Bonje Model 4 eGFR <90
	Coefficient (95%CI);	Coefficient (95%CI);	Coefficient (95%CI);	Coefficient (95%CI);
Age [c]				
Per 10-year increase	-8.84(-9.53, -8.14)	-9.00 (-9.79, -8.22)	3.04 (2.33, 4.13)	3.09 (2.12, 4.59)
Sex [d]				
Male	0.22 (-1.66, 2.12)	-0.10 (-2.45, 2.23)	0.78 (0.32, 1.86)	1.55 (0.37, 8.28)
Female	Ref	Ref	Ref	Ref
Education (years)				
≤5	4.35 (0.61, 8.09)	4.19 (0.39, 7.99)	0.77 (0.17, 3.33)	0.92 (0.17, 4.58)
>5≤10	1.5(-0.46, 3.55)	1.65 (-0.38, 3.69)	1.00 (0.35, 2.96)	1.04 (0.33, 3.40)
>10	Ref	Ref	Ref	Ref
Occupation				
Agricultural worker	-0.99 (-3.05, 1.07)	-1.04 (-3.14, 1.04)	0.48 (0.18, 1.31)	0.45 (0.17, 1.27)
Non-agricultural worker	Ref	Ref	Ref	Ref
Household monthly income (MK)[e]				
Unknown	-3.86(-11.05, 3.31)	-4.46 (-11.76, 2.82)	2.31(0.08, 27.04)	3.18 (0.11, 37.04)
MK 0 ≤20,000	Ref	Ref	Ref	Ref
MK >20,000	0.05 (-1.91, 2.03)	0.35 (-1.67, 2.37)	1.18 (0.47, 2.93)	1.02 (0.37, 2.73)
BMI (kg/m^2)				
5kg/m^2 increase	-0.71 (-2.05, 0.62)	-0.62 (-1.97, 0.73)	1.76 (1.05, 3.01)	1.89 (1.06, 3.81)
Fat Free Mass (kg)				
(Per 5 kg/m2 unit increase)	-0.02 (-0.08, 0.02)	-0.01 (-0.07, 0.04)	0.99 (0.00, 1.05)	0.92 (0.58, 1.01)
Healthy lifestyle choices				
Non-smoker or alcohol drinker	-1.53 (-4.03, 0.96)	-1.40 (-3.95, 1.13)	1.47 (0.45, 5.13)	1.52 (0.43, 5.87)
Smoker and alcohol drinker	Ref	Ref	Ref	Ref
Regular meat-eater				
Yes	0.00 (-2.03, 2.04)	-0.10 (-1.77, 2.77)	1.31 (0.53, 3.52)	1.05 (0.41, 2.93)
No	Ref	Ref	Ref	Ref

[a] Percentages in columns; [b] percentages in rows; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30mg/mmol; [c] Exchange rate (MK to USD) 0.001 at time of questionnaire

Appendix 25. Supplementary table showing linear models for Area 25

Table 1. Linear regression models, showing both minimally and fully adjusted models, Area 25 (n=243)

Variable	Area 25 Model 1 eGFR Coefficient (95%CI);	Area 25 Model 2 eGFR Coefficient (95%CI);
Age [c]		
Per 10-year increase	-8.26 (-9.91, -6.67)	-8.14 (-10.61, -6.23)
Sex [d]		
Male	-0.49 (-3.97, 2.98)	-1.62 (-5.29, 2.05)
Female	Ref	Ref
Education (years)		
≤5	5.53 (-2.55, 13.62)	5.59 (-2.63, 13.82)
>5≤10	-0.06 (-4.83, 4.570)	-0.62 (-5.49, 4.24)
>10	Ref	Ref
Occupation		
Agricultural worker	-0.49(-15.09, 14.10)	0.99 (-13.76, 15.75)
Non-agricultural worker	Ref	Ref
Household monthly income (MK)[e]		
Unknown	-0.11 (-10.29, 10.05)	-0.83 (-11.10, 9.43)
MK 0 ≤20,000	Ref	Ref
MK >20,000	-2.24 (-6.87, 2.38)	-1.73 (-6.45, 2.98)
BMI (kg/m^2)		
5kg/m^2 increase	-1.61 (-3.53, 0.29)	-1.58 (-3.54, 0.36)
Fat Free Mass (kg)		
(Per 5 kg/m2 unit increase)	-0.02 (-0.11, 0.07)	-0.02 (-0.11, 0.06)
Healthy lifestyle choices		
Non-smoker or alcohol drinker	-1.52(-5.46, 2.42)	-1.84 (-5.83, 2.14)
Smoker and alcohol drinker	Ref	Ref
Regular meat-eater		
Yes	2.62 (-7.90, 2.66)	-1.95 (-7.38, 3.46)
No	Ref	Ref

[a] Exchange rate (MK to USD) 0.001 at time of questionnaire; Hypertension = systolic bp ≥140 mm Hg, or diastolic bp ≥90 mm Hg; Diabetes = fasting glucose >=7mg/l; Proteinuria = ACR >=30

Appendix 26: Nested environmental study ethics approval

Telephone: +265 789 400
Facsimile: +265 789 431

All Communications should be addressed to:
The Secretary for Health and Population



In reply please quote No. MED/4/36c
MINISTRY OF HEALTH AND POPULATION
P.O. BOX 30377
LILONGWE 3
MALAWI

6th April, 2018

Amelia Crampin
MEIRU

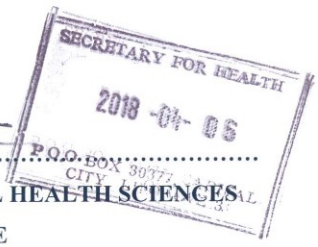
Dear Sir/Madam,

Re: Protocol # 17/09/1912: Cross Sectional Study on Prevalence and Risk Factors of Chronic Kidney Disease of Unknown Cause (CKDu) in Malawi Comparing Urban and Rural Populations

Thank you for the above titled proposal for amendments that you submitted to the National Health Sciences Research Committee (NHSRC) for review.

Please be advised that the NHSRC **reviewed and approved** the adding of a questionnaire for the study

Kind regards from the Secretariat.



.....

FOR: CHAIRMAN, NATIONAL HEALTH SCIENCES RESEARCH COMMITTEE

Executive Committee: *Dr B. Chilima (Chairperson), Dr B. Ngwira (Vice-Chairperson)*
Registered with the USA Office for Human Research Protections (OHRP) as an International IRB
IRB Number IRB00003905 FWA00005976

Appendix 27. Nested environmental questionnaire information sheet

KDU Environmental Questionnaire Information sheet

This questionnaire has been written by a PhD researcher working on the KDU study who would like to understand your daily activities and/or employment, and your general lifestyle.
Mafunsowa alembedwa ndi mkulu wa kafukufuku wokhudza impso amene akufuna kumvetsetsa za ntchito zomwe mumagwira tsiku ndi tsiku komanso zokhudza mmene mumakhalira pa moyo wanu.
This project is additional to the core KDU study, looking at kidney health from an environmental and geographical perspective. No personal or health questions will be asked in this questionnaire.
Kafukufukuyu ndi owonjezera pa kafukufuku yemwe alipo kale wa impso, kufuna kuwona za impso maka ku mbali ya zomwe zakuzungulirani ndi za nyengo komwe mumakhala. Simudzafunsidwa zokhudza inu kapena umoyo wanu mu chikalata ichi.
You will be asked if you wish to take part in this questionnaire and to sign a consent form. The field interviewer will conduct the questionnaire after they have completed the core KDU questionnaire. This questionnaire will take no more than 30 minutes to complete.
Mudzafunsidwa ngati muli olola kutengapo mbali pa kafukufukuyu ndipo mudzayenera kuonetsa kuvomereza kwanu posayinira chikalata. Ogwira ntchitoyi adzakufunsani mafunso awa atamaliza kukufunsani zokhudza kafukufuku wa impso yemwe alipo kale.
Mafunso awa adzafunsidwa pa mphindi zosapitilira makumi atatu.
The document is comprised of four sections:
Chikalatachi chiri ndi magawo okwana anayi:
Part 1: Household information
Gawo loyamba: lokhudzana ndi panyumba
Time living at current residence
Nthawi yomwe mwakhalako kumene mukukhala panopa
Part 2: Occupation
Gawo lachiwiri: Ntchito
Questions about your employment or main activities throughout a normal day
Mafunso okhudza ntchito zomwe mumagwira pa tsiku
Part 3: Agricultural questionnaire
Gawo lachitatu: mafunso okhudza ulimi
For those who farm (subsistence, small-holder commercial, tending to gardens), this part aims to capture farming practices throughout the year
Kwa omwe amalima (zakudya, zogulitsa pa mulingo wochepa, kusamalira minda yaing'ono), mbali iyi, ndi yofuna kudziwa za njira za malimidwe mchaka chonse.
Part 4: Lifestyle
Gawo lachinayi: makhalidwe
Diet and general lifestyle questions (such as drinking water sources, cooking practices)
Mafunso a zakudya ndi mmene mumakhalira (monga, malo otunga madzi okumwa, komanso mmene mumaphikira)

Results will be published in journal papers and a thesis, but no personal identifiers will be used – your identity will remain anonymous.
Zotsatira zidzasindikizidwa mmapepala, koma sipadzakhala chirichonse chodziwitsa za inu – simudzadziwika.
Information will be communicated back to the community along with the core KDU information, post-study completion.
Zotsatirazi pamodzi ndi za kafukufuku yemwe alipo kale wa impso zidzaperekedwa kwa anthu omwe atenga mbali mu akafukufuku onse awiri.

Appendix 28. Nested environmental study consent form

MALAWI EPIDEMIOLOGY AND INTERVENTION RESEARCH UNIT (MEIRU)

CHRONIC KIDNEY DISEASE OF UNKNOWN ORIGIN – ENVIRONMENTAL AND LIFESTYLE QUESTIONNAIRE CONSENT FORM – ENGLISH

Your Consent: Please answer the questions below to show that you agree to take part in the environmental and lifestyle questionnaire.

You will also be required to write down your name together with today's date at the bottom of the page. The interviewer will also counter sign. A copy of this consent will be kept by the interviewer and another will be given to you to keep.

Study ID	*Study ID*		STID

Date completed	*Date*		INTDATE

1	I have read or have been read the information concerning the environmental and lifestyle study and I have been given the information sheet. I have been given the chance to ask all the questions to help me understand why this study is being done. I know that my participation is voluntary. I understand my personal details such as name, phone number and household will not be revealed to people outside the project. I know that the final results will not contain any names	1=Yes CONSENT A 0= No
2	I agree to take part in this study and to answer the questionnaire as fully and honestly as possible	1=Yes CONSENTB 0=No
3	I understand that the information I provide may be archived by the project I understand that the information I provide may be used in publications, reports, web pages, and other research outputs in an anonymised form. I understand that other researchers may have access to this data only in a securely anonymised form.	1=Yes CONSENTC 0=No

Name of participant		
First name (s)		
Surname		

Appendix 29. Environmental study questionnaire

Interview date:	d	d	m	m	y	y	y	y

Staff code

STID: Yellow label Stick here

CNO: Orange label Stick here

Part 1: Household information Questionnaire

1.1 How long have you lived in your current house?

weeks		months		years	

Part 2: Occupation Questionnaire

Question aim: to capture what participant does with their day – capture activities
If participant says they are unemployed or have no work to do - ask what activities they do every day – for example, if they stay at home and do housework, or tend to their gardens, please probe them to get a good idea of this.
Those at home or "unemployed" could still be conducting strenuous manual labour, so very important to get this information
If participant does many of these activities per day, please ask them what the MAIN activity of their day is
If they split the day into 2 activities (office work and tending crops) then please choose both –

2.1 What is your current occupation/daily activity?

Occupational category	*Tick as applicable*	
Street trader/vendor **Answer questions: 2.2-2.4 THEN go to question 2.7**		1
Shop trader/vendor **Answer questions: 2.2-2.4 THEN go to question 2.7**		2
Manufacturer (Detail below what you produce) **Answer questions: 2.2-2.4 THEN go to question 2.7**		3
Office worker (Detail the office location and industry you work for) _____ **Answer questions: 2.2-2.4 THEN go to question 2.5**		4
Construction worker **Answer questions: 2.2-2.4 THEN go to question 2.6**		5
Commercial agriculture worker **Answer questions: 2.2-2.4 THEN go to question 2.7**		6
Casual labourer **Answer questions: 2.2-2.4 THEN go to question 2.7**		7
Subsistence farming/fishing **Answer questions: 2.2-2.4 THEN go to question 2.7**		8
Carrying or transporting items or people **Answer questions: 2.2-2.4 THEN go to question 2.6**		9
Homemaker (household/land activities) **Answer questions: 2.2-2.4 THEN go to question 2.7**		10
Full-time student		11

Go to part 4		
Other **(please detail)** **Answer questions: 2.2-2.4 THEN go to question 2.7**		999

Question aim: identify how long participant has been in their employment, or doing the activities they discuss in question 2.1

2.2 How long have you been conducting these work/activities that you selected above?

months | years

Question aim: capture the start and end of typical day
This can be answered as a time (06:00)
or
as a description: before dawn, after dawn (and so on)

2.3 **At what time do you commence, and finish your daily activities?**
Please detail below:

Question aim: quantify question 2.3 into a number of hours

If participant does now know how many hours a day they work, ask probing questions such as

"do you work from dawn to dark?"

"Do you work or conduct your activities before the sunrise?"

2.4 As selected above (question 2.3) how many hours do you conduct work/labouring/household activities per day		
Hours worked per day	***Tick as applicable***	
None		1
0-2		2
2-4		3
4-6		4
6-8		5
8-10		6
>10		7
Other (please detail)		999

This question is for those participants working in an office space only

"Other" – use this box if participant says they use a cooling method which is not listed below

2.5 Does your office have any of the following facilities		
Office facilities	***Tick as applicable***	
Ceiling fan		1
Portable fan		2
Air conditioning		3

Other (please detail)	999

Question aim: to understand if participant uses physical activity to transport or carry items, or, if they use motorized vehicles (non-physical activity)

Please ask which mode of carrying or transporting the participant uses and write in the "tick as applicable" box

2.6 Which type of transport do you use for your carrying/transporting activities? (you can tick more than one if this applies to you		
Mode of transport for carrying items	***Tick as applicable***	
Physical (for example walking or bicycle taxi)		1
Non-physical (for example car, bus, motor bike)		2
Other (please detail)		999

This question assumes a 7-day week (Monday-Sunday)

2.7 **How many days per week do you conduct your activities?**

	Tick as applicable
1-2	1
2-3	2
3-5	3
5-7	4
Don't know	000

Question aim: to capture how often the participant does this job or carries out these activities

Are they seasonal farmers? Subsistence/commercial famers/fishermen all year? Do they work in their office or shop seasonally? (and so on)

2.8 **Are your activities:**

	Tick as applicable
Seasonal	1
All year	2
Other:	999

Question aim: does the participant stop for a period to rest/eat/drink whilst conducting their work or activities?

If they describe their day and when they take breaks, please write this in the "other" box

2.9 Do you rest in the day between conducting your daily activities?		
Frequency of rest during the working day	***Tick as applicable***	
Hourly		1
Every 2-3 hours		2
Every 4-6 hours		3
Every 6-8 hours		4
Never		5
Other (please detail)		999

Question aim: when participant stops to rest/eat/drink, for how long do they stop?

Probe – if participant does not know the timings, then ask how long it takes for them to eat lunch/drink their water and ask if that is how long they stop for and add in the detail in "other" box

2.10 How long are your rest breaks in general?		
Frequency of rest during the working day	***Tick as applicable***	
5-10 minutes		1
10-15 minutes		2
15-20 minutes		3
More than 20 minutes		4
Other (please detail)		999

Question aim: capture any ill-feelings as a result of participants job/daily activities

If participant experiences more than one, please tick as applicable

If participant experiences other symptoms not listed, please write in the "other" box

2.11 Do you experience any of the following symptoms as a result of your daily activities:			
	Yes	***No***	
Raised heart rate			1
Sweat			2
Tired			3
Weakness			4
Feeling faint			5
Muscle cramps			6
Thirst			7
Nausea			8
Other (detail below if applicable to you)			999

Part 3 aim: Intended for those who **tend to crops and land**

This can be for those tending to crops on: commercial farms, small land plots, gardens (large and small)

Part 3: Agricultural Questionnaire

Question aim: location of where participant tends crops

This can be for **all participants** who tend to crops (can be home address if applicable)

3.1 Farm location (address)

Question aim: To capture the crops the participant tends to (on farm/garden/ other plots)

If there is more than once crop tended to, please select in the table

3.2 Crop(s) under production on your land/garden/farm

	3.2.1 Produced	3.2.2. Purpose	3.2.3. Method of production

Crop type	Yes (1)	Subsistence (1)	Small-holder (2)	Commercial (3)	Manually watered (1)	Rain fed (2)	Large-scale irrigation (3)
Cereals							
Wheat							
Rice							
Maize							
Sorghum							
Rape							
Other cereals (please list)							
Pulses							
Beans (type)							
Other pulses (please list):							
Fruits							
Tomato							
Banana							
Pineapple							
Guava							
Mango							
Papaya							
Lemon							
Other fruits (please list)							
Root vegetables							
Irish Potato							
Sweet potato							
Groundnuts							
Onion							
Pumpkin							
Other (please list)							
Vegetables							
Cabbage							
Wild seed potato leaf							
Spinach							
Other (please list)							
Plant or tree crops							
Tobacco							
Sugarcane							
Cassava							
Cotton							
Cocoa							
Tea							
A1. Coffee							
B1. Other plant/tree crops (please list)							
Dairy produce							
C1. Milk							
D1. Other dairy produce							
Livestock							
E1. Dairy Cattle							
F1. Goat							
G1. Fowl							
H1. Chicken							
I1. Pigs							
J1. Other livestock (please list)							
G1. Other crops (please list)							

Question aim: How big is the farm or plot of land that the participant tends to?

3.3 How many hectares are under production on the farm or garden where you tend crops?		
	Tick as applicable	
<1 hectare		1
1-5 hectares		2
5-10 hectares		3
10-15 hectares		4

>15 hectares		5
Don't know		000

Question aim: In general, does participant know if pesticides are applied to crops they tend to

This can be across a large commercial farm, or on a small garden plot

3.4 Are pesticides and/or herbicides applied to the crops on the land/farm?		
	Tick as applicable	
Yes		1
No		2

3.5 Do you apply pesticides to the crops yourself?		
	Tick as applicable	
Yes		1
No		2

If you answered "yes" to question 3.5, please answer questions 3.6-3.13. **Otherwise go to Part 4.**

Question aim: Understand pesticide handling practices for participants who directly apply pesticides
Please ask each question in **bold** below
Please note – this is applicable for **all** participants who tend to crops – from a large commercial farm, to a small garden plot

Please Tick options applicable to your pesticide/herbicide use practices below

	Tick as applicable
3.6.1 Where do you mix your pesticides?	
Near a community water source (river, canal, well)	1
In the field (farm)	2
At your residence	3
Other (please detail)	999
3.6.2. How do you mix your pesticides/herbicides?	
Bare hands	1
Hands, wearing gloves	2
With a stick, and bare hands	3
With a stick, wearing gloves	4
Other (please detail)	999
3.6.3. Which devices do you use for mixing pesticides/herbicides?	
Container (bucket, drum etc.)	1
Other (please detail)	999
3.6.4. Do you use any personal protective equipment (PPE) during application of pesticides/herbicides?	
Wearing normal clothing	1
Overalls	2
Face mask	3
Hat	4
Boots	5
Barefoot	6
Bathing after application	7
Other (please detail)	999
3.6.5. Disposal of empty pesticide containers	
Dump on the field	1
Throw into irrigation canals/rivers/streams	2
Bury on farm	3

Burn on farm	4
Do not dispose – keep for domestic purposes	5
Do not dispose – collect and sell them	6
Other (please detail)	999
3.6.6. Where do you store your pesticides/herbicides?	
At your residence (outside)	1
Inside your house	2
On your farm	3
Other (please detail)	999
3.6.7. How do you store your pesticides/herbicides?	
Plastic bag	1
Container/drum (please detail if open or sealed)	2
Storage facility (cupboard etc.)	3
Other (please detail)	999

Question aim: Understand method of direct pesticide/herbicide application

3.7 How do you apply pesticides and/or herbicides to crops?		
Pesticide/herbicide application method	***Tick as applicable***	
Knap-sack spray (manual)		1
Knap-sack spray (motorized)		2
Pour directly from container		3
Other (please detail)		999

3.8 How often do you apply pesticides/herbicides to crops you tend?		
Frequency of pesticide/herbicide application	***Tick as applicable***	
Every day		1
Every 2 days		2
Every 3 days		3
Every 4 days		4
Every week		5
Every 2 weeks		6
Every month		7
Other (please detail)		999

Question aim: Daily timing of application
This can be multiple depending on the crop/season
Probe: If participant does not know timing, ask in a different way – for example:
"Do you apply before dawn" "Do you apply after sunset" and so on…

3.9 What time of the day do you apply pesticides/herbicides?		
Frequency of crop application	***Tick as applicable***	
Morning		1
Midday		2
Late afternoon		3
Entire day		4
Other (please detail)		999

Question aim: Annual pattern of pesticide/herbicide application
Does participant apply seasonally? Does participant apply in selected months of the year?

3.10 What time of the year do you apply pesticides/herbicides?		
Pesticide/herbicide application throughout the year	***Tick as applicable***	
Rainy season		1

Dry season		2
Combination		3
Month(s) (please detail)		4

Question aim: Does the participant know which pesticides/herbicides they are using?
Probe: how is the pesticide/herbicide packaged? In a bag? In a bucket? In a drum? And so on…
Probe: Does the participant know where the pesticide/herbicide was purchased from? If so, please write this in the "other" box (this can be shop name, company and so on)

3.11 Do the pesticides/herbicides you apply have labels on the container?		
Pesticide/herbicide labelling	***Tick as applicable***	
Label with name of pesticide/herbicide		1
Label showing chemical ingredients of the pesticide/herbicide		2
No label		3
Other (please detail)		999

Question aim: Understand if participant knows either the commercial (label) name, or the active ingredient in the pesticides/herbicides they use
Probe: if possible, ask to see the pesticide packaging they use – note the name down/take photo

3.12 If you know the name or active ingredient of the pesticide(s)/herbicide(s) you use, list them below

Name
Name:
Name:

Question aim: does the participant wash themselves directly after they have finished applying pesticides/herbicides? This can include: washing hands, arms, face, feet, and so on
If participant tells you the method of cleaning (bath, bucket, towel etc) please write this down next to the "other" box

3.13 Do you wash your body after applying pesticides/herbicides?		
	Tick as applicable	
Yes		1
No		2
Sometimes		3
Other (please detail)		999

Part 4: Lifestyle Questionnaire

Question aim: Does the participant have a main water source they drink water from during the year, and a second water source for other times of the year?
(for example, do they drink from a tap during dry season and borehole during wet?)
Primary water source definition: Main water source for the year
Secondary water source definition: Water source used when the main source is not working/too dry
If participant has one water source all year, leave secondary source boxes empty

4.1 **What are your main drinking water sources?**

Main drinking water source(s)	**4.1.1. Primary Source**	**4.1.2. Location (area/address)**	**4.1.3. Secondary Source**	**4.1.4. Location (area/address)**	
Piped into dwelling					1
Piped into your yard/plot					2
Open well in your yard/plot					3
Open public well					4
Protected well in your yard/plot					5

Protected public well (water kiosk)	6
Borehole	7
Spring	8
River	9
Stream	10
Pond	11
Lake	12
Dam	13
Reservoir	14
Rainwater	15
Bottled water	16
Deep well	17
Shallow well	18
Other (please detail)	999

Question aim: To understand time of year participant uses their water source(s)

If certain months are provided, then please detail in the "other" box

4.2 D**o you use the same primary drinking water source all year long**

Tick as applicable	
Yes	1
No, Only in the rainy season	2
No, Only in the dry season	3
Other (please detail)	999

4.3 D**o you use the same secondary drinking water source all year long**

Tick as applicable	
All year	1
Only in the rainy season	2
Only in the dry season	3
Other (please detail)	999

Question aim: Using a standard-sized cup as a prompt, ask participants how many they will fill with water and drink per day

4.4 **How much water do you drink per day?**

Daily water consumption	***Tick as applicable***
<1 cup	1
1-2 cups	2
2-3 cups	3
3-4 cups	4
4-5 cups	5
More than 5 cups	6
Don't know	000
Other (please detail)	999

Question aim: Understand if the participant applies purification methods to their drinking water before consuming

Probe: if participant needs prompting, give examples such as: boiling, filtering and so on

4.5 Do you treat your drinking water before consuming it?		
	Tick as applicable	

Yes		1
No		2

Question aim: Method used for water treatment
This is not limited to filtering and boiling, please tell participant they can detail any other methods they might use for treatment (if any)

4.6 If you selected "yes", please state which method(s) you use

If you selected "no", proceed to question 4.7

Fuel for cooking	***Tick as applicable***
Water filter	1
Boiling	2
Other (please detail)	999

4.7 How many times a week do you eat:

Foodstuff	1-2 times a week (1)	2-3 times a week (2)	3-4 times a week (3)	>4 times a week (4)	Never (5)
4.7.1 Fish					
4.7.2 Meat (any)					
4.7.3 Offal					
4.7.4 Maize					
4.7.5 Rice					
4.7.6 Cassava					
4.7.7 Vegetables (any)					
4.7.8 Pulses					

Appendix 30. Shipping invoice whole blood samples Malawi to London

MEIRU
Malawi Epidemiology and Intervention Research Unit

CUSTOMS INVOICE

Date: 09th May, 2019

Shipper: LOUIS G. BANDA
MEIRU
CHSU Campus
Area 3,
LILONGWE
MALAWI
Email: louis.banda@kpsmw.org

Phone number: +265 888 325 887 / +265 992 565 755

Consignee: Dr Scott Young,
Associate Professor and Reader in Environmental Science,
University of Nottingham
Gateway Building
Sutton Bonington Campus
Loughborough
Leicestershire

Contact number:

Contents*: **Biological Substance, Category B**
Packed in accordance with P.I.650, UN3373

Total Quantity: ***5 Freezer boxes (81 Well)*** **(13 x 13 x 4.7 cm)**

Sample type: EDTA Whole blood = *163 X 1500uL Vials*

Human blood/blood products specifically:

- To be used for laboratory analysis only.
- Exclusively for research purposes in laboratory. Not for sale
- Material obtained directly from humans that have not been inoculated with or exposed to any livestock or poultry disease agent
- Material is not of tissue culture origin and not recombinant
- Material is not an actual zoonotic pathogen
- Material is not a potential zoonotic pathogen

***No Commercial Value - For Customs Purposes, the Value = $10.00 USD**

Appendix 31. Physicochemical parameters

Temperature

Cool water is generally more palatable than warm water, and temperature impacts the acceptability of several other inorganic constituents and chemical contaminants. High water temperature may increase problems related to taste, odour and colour [127].

pH

pH is a measure of how acidic or basic water is, and ranges from 0 to 14, with 7 being neutral. A pH of <7 indicates acidity, whereas a pH of > 7 indicates a base. pH is an important indicator of the chemical concentration of water and determines the solubility and biological availability of chemical constituents such as nutrients and heavy metals. In the case of heavy metals, the degree to which they are soluble determines their toxicity, and thus potential human health impacts. The optimum pH ranges from 6.5–8.5[24, 127].

Dissolved Oxygen [DO]

DO refers to the level of free, non-compound oxygen present in water, and is influenced by source, temperature, and chemical or biological processes. Previous research has shown that DO plays an important role in determining the contents of heavy metal in sediments, with more heavy metals being released from sediments in an anoxic environment and adsorbed from the water in an aerobic environment[335].

Electrical Conductivity

The EC is a measure of the dissolved ionic component in water and gives an indication of the amount of total dissolved substitution in water. An increase in ion concentration (dissolved metals) enhances the EC, and generally, the level of dissolved solids in water determines the EC[78], and can impart substantial water palatability issues, which result in indirect impacts on health due to limited water consumption by the consumer.

Total dissolved solids [TDS]

TDS comprise inorganic salts (anions and cations of calcium, magnesium, sodium, chlorides and sulfates) and trace organic matter that are dissolved in water. Sulfates are hypothesised to catalyse a release of heavy metals into a water system in the presence of Fe, especially in ground water sources[362], and chlorides are known to react with metal ions (especially in metal pipes), forming soluble salts, thus increasing levels of metals in water[127]. Calcium and magnesium are indicative of water hardness [$CaCO_3$], which is not known to have a direct adverse health effect, however there are some links to cardiovascular disease and kidney stones[363]. $CaCO_3$ will be included in further analyses as it is a heavily studied in Sri Lanka, and although the causal link cannot yet be explained consistently, it is strongly hypothesised to be a putative factor (in synergy with other metals) for CKDu in the region [90, 129, 364].

Turbidity

Turbidity describes the cloudiness of water caused by suspended particles chemical precipitates, organic particles and organisms. Turbid water can cause dissolved metals to attach to suspended particles in water, meaning that an increase in turbidity can often indicate potential pollution [365].

Appendix 32.

The University of Nottingham
UNITED KINGDOM · CHINA • MALAYSIA

Faculty of Science

School of Biosciences

Division of Agricultural and Environmental Science

Sutton Bonington Campus
Sutton Bonington
Loughborough

To whom it may concern

This is to confirm that Sophie Hamilton is a postgraduate research student at Imperial College, London. As part of her research project, she needs to import small samples of water and other environmental materials into the UK, for chemical analysis, under our DEFRA Plant Health licence number 105352/206483/3.
All samples will be destroyed by incineration following chemical analysis by Ion Chromatography and Inductively Coupled Plasma Mass Spectroscopy.

Yours faithfully,



Dr Scott Young
Associate Professor and Reader in Environmental Science,
University of Nottingham.
Tel. 0115 9516256
scott.young@nottingham.ac.uk

Appendix 33. UoN DEFRA

1. Name and address of consignor/Plant protection organization of the country of origin	Letter of Authority for the introduction and/or movement of harmful organisms, plants, plant products and other objects for trial or scientific purposes and for work on varietal selections (issued under Directive 2008/61/EC)
2. Name and address of person responsible for the approved activities Mr J. Corrie Division of Agriculture and Environmental Science University of Nottingham, School of Biosciences, Gateway Building, Sutton Bonington Campus Loughborough, Leicester LE12 5RD	3. Name of the responsible official body of the Member State of issue Department for Environment, Food and Rural Affairs (Defra) Welsh Government
4. Address and description of the specific site or sites for quarantine containment Project 1; Gateway Building Laboratory: A14., B06. & Room A22a (Soil store) at the above address. Project 2; Laboratory B05, BO6, B10, A12, A13 (Analytical laboratories), Storage and preparation laboratory- A14, and locked unit within soil store A22a at the above address.	5. Place of origin (documentary evidence attached for material originating in a third country) As per description in box 8
	6. Plant passport number: or Phytosanitary certificate number:
7. Declared point of entry for material introduced from a third country Great Britain	
8. Scientific name(s) of the material, including the harmful organisms concerned:	9. Quantity of material As Required
Project 1 Soil and associated plant material (dried and ground) From: All countries Project 2 Soil and organic material (dried plant roots and leaves) From: All countries	
10. Type of material Licence to import, move and keep prohibited soil (other than for chemical and physical analysis)	
11. Additional declaration This material is introduced into/moved within (1) the Community under Commission Directive 2008/61/EC and Defra plant health licence no 105352/206483/3	
12. Additional information Valid from 01 January 2018 to 31 December 2018	

13. Endorsement by the responsible official body of the Member State of origin of the material Place of endorsement: Date: Signature of authorized officer: Name in BLOCK LETTERS	14. Stamp of the responsible official body of issue Place of issue: YORK Date: 09 November 2017  Signature of authorized officer:  Name in BLOCK LETTERS ZARA GOODWIN

APPENDIX 34: Inductively coupled plasma mass spectrometry (ICP-MS) and Ion chromatography (IC) methods used in the Nottingham laboratory to test for heavy metals in potable water sources

Inductively coupled plasma mass spectrometry

The instrument was run employing three operational modes, including (i) a collision-cell (Q cell) using He with kinetic energy discrimination (He-cell) to remove polyatomic interferences, (ii) standard mode (STD) in which the collision cell is evacuated and (iii) hydrogen mode (H_2-cell) in which H_2 gas is used as the cell gas. Samples were introduced from an autosampler (Cetac ASX-520) incorporating an ASXpress™ rapid uptake module through a perfluoroalkoxy (PFA) Microflow PFA-ST nebuliser (Thermo Fisher Scientific, Bremen, Germany). Internal standards were introduced to the sample stream on a separate line via the ASXpress unit and included Ge (10 µg L^{-1}), Rh (10 µg L^{-1}) and Ir (5 µg L^{-1}) in 2% trace analysis grade (Fisher Scientific, UK) HNO_3. External multi-element calibration standards (Claritas-PPT grade CLMS-2 from SPEX Certiprep Inc., Metuchen, NJ, USA) included Ag, Al, As, Ba, Be, Cd, Ca, Co, Cr, Cs, Cu, Fe, K, Li, Mg, Mn, Mo, Na, Ni, P, Pb, Rb, S, Se, Sr, Tl, U, V and Zn, in the range 0 – 100 µg L^{-1} (0, 20, 40, 100 µg L^{-1}). A bespoke external multi-element calibration solution (PlasmaCAL, SCP Science, France) was used to create Ca, Mg, Na and K standards in the range 0-30 mg L^{-1}. Phosphorus, boron and sulphur calibration utilized in-house standard solutions (KH_2PO_4, K_2SO_4 and H_3BO_3). In-sample switching was used to measure B and P in STD mode, Se in H_2-cell mode and all other elements in He-cell mode. Peak dwell times were 10 mS for most elements with 300 scans per sample. Sample processing was undertaken using Qtegra™ software (Thermo-Fisher Scientific) utilizing external cross-calibration between pulse-counting and analogue detector modes when required.

Ion chromatography

Ion chromatography (IC) is the separation and quantitative analysis of anions and cations in an ionic solution using the ion exchange method of liquid chromatography (LC). The process separates the different ions within a sample, and the amount of anion/cation is measured by the change in conductivity as the species ass through the detector. The Thermo Scientific Dionex ICS-1100 Ion Chromatography System instrument was used to conduct this analysis, with (a) IonPac AG14A Carbonate Eluent Anion-Exchange Column guard column, and (b) IonPac AS14A Carbonate Eluent Anion-Exchange Column analytic column. The sample was passed through eluent (Na2C03 (3.5mM) and NaHCO3 (1.0mM)) at a flow rate of 1.4mL/min at 30°C.

The standards used to calibrate the IC are listed below:

Name	Total Volume (mL)	F		Cl		$N\text{-}NO_3$		$P\text{-}PO_4$		$S\text{-}SO_4$	
		ppm	mL Stock	ppm	mL Stock	ppm	mL Stock	ppm	mL Stock	ppm	mL Stock
STD 0	ELUENT										
STD 1	100	1	0.1	1	0.1	1	0.1	1	0.1	1	0.1
STD 2	100	2	0.2	2	0.2	2	0.2	2	0.2	2	0.2
STD 3	100	3	0.3	3	0.3	3	0.3	3	0.3	3	0.3
STD 4	100	4	0.4	4	0.4	4	0.4	4	0.4	4	0.4
STD 5	100	5	0.5	5	0.5	5	0.5	5	0.5	5	0.5
STD 10	100	10	1	10	1	10	1	10	1	10	1
STD 20	100	20	2	20	2	20	2	20	2	20	2

Appendix 35. Malawi heat index data distributions

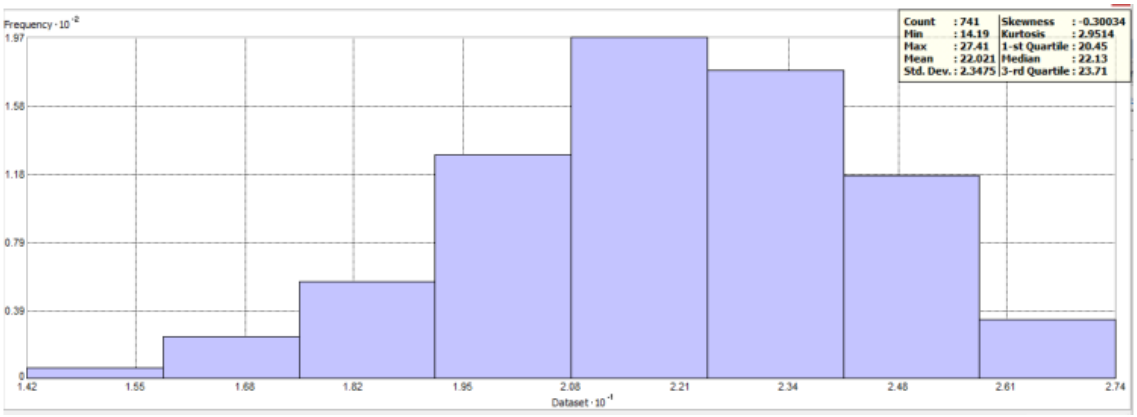


Figure 1. Histogram of heat index data for Malawi

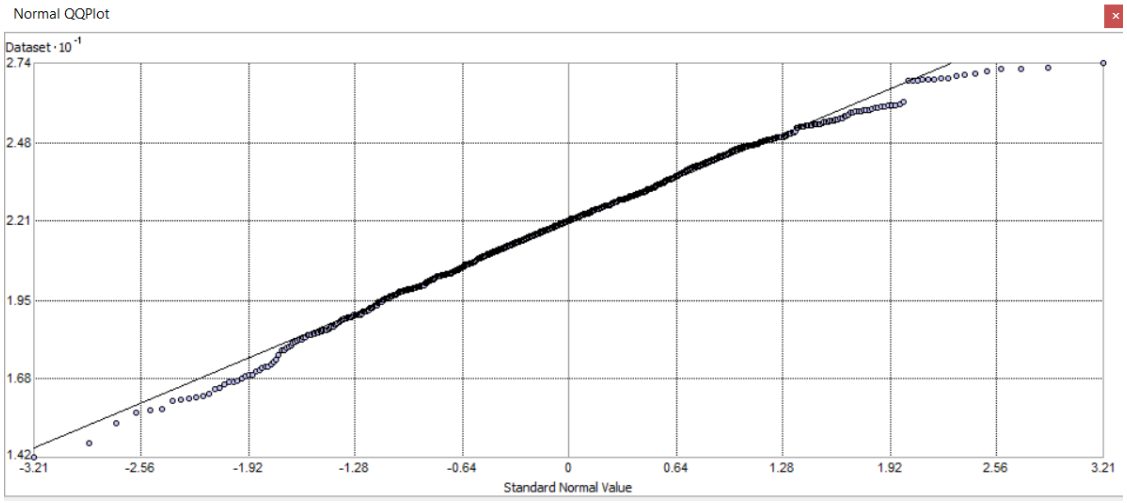
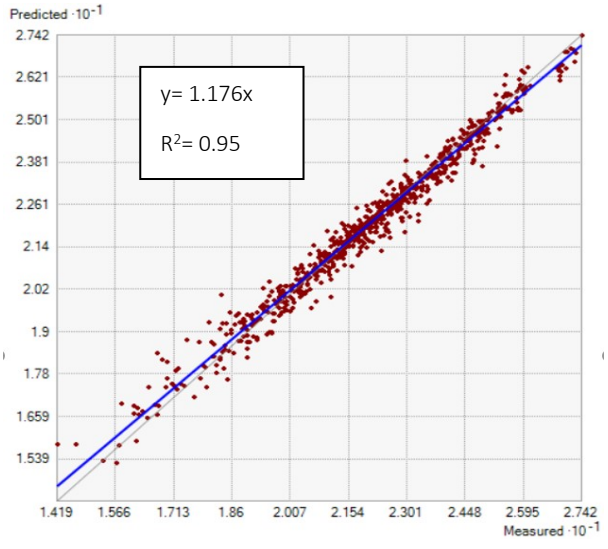


Figure 2. QQ plot of heat index data for Malawi

Appendix 36. EBK heat index model fit for Malawi



Appendix 37. Sensitivity analyses for heat index across Malawi

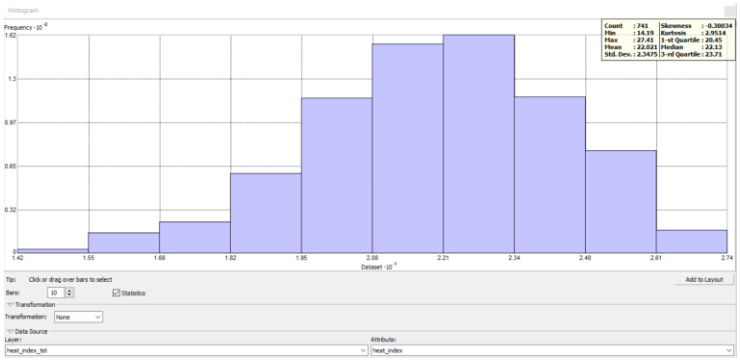


Figure 1. Histogram of data distribution of restricted heat index data across Malawi

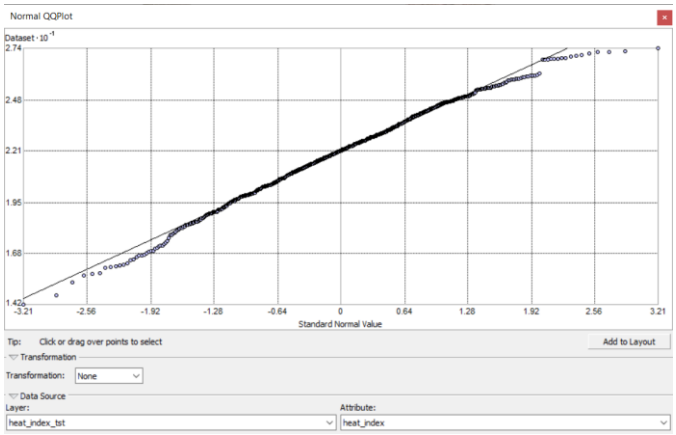


Figure 1.QQ plot of restricted heat index data across Malawi

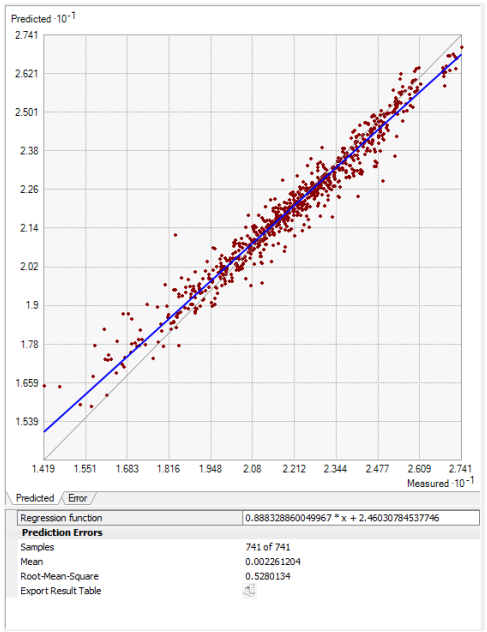


Figure 3. Model fit for inverse distance weighting data across Malawi

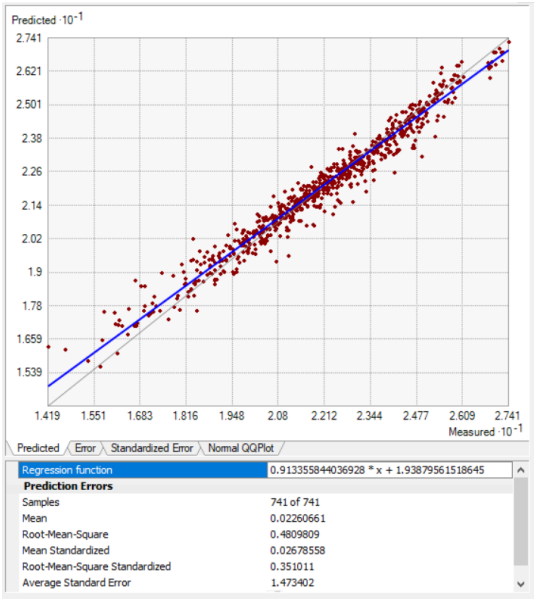


Figure 4. Model fit for kriging data across Malawi