Accepted Manuscript

Investigating relations between environmental toxins in northern Irish soils and streams and Chronic Kidney Disease prevalence

Chloe E. Jackson, Jennifer M. McKinley, Ulrich Ofterdinger, Damian Fogarty, Peter M. Atkinson, Sherry Palmer

Investigating Relations between Environmental Toxins in Northern Irish Soils and Streams and Chronic Kidney Disease Prevalence

Chloe E. Jackson, Jennifer M. McKinley[1], Ulrich Ofterdinger[2], Damian Fogarty[3], Peter M. Atkinson[4,1,5], Sherry Palmer[6]

1. School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, N. Ireland.

2. EERC, School of Planning, Architecture and Civil Engineering, Queen's University Belfast, N. Ireland

3. Belfast Health Trust, Lisburn Road, Belfast, Northern Ireland

4. Lancaster University, Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, United Kingdom

5. Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK

6. EERC, School of Planning, Architecture and Civil Engineering, Queen's University Belfast, N. Ireland

Abstract

The unknown aetiology of Chronic Kidney Disease (CKD) has attracted recent attention as a result of the increasing global prevalence and recent reviews of occupational and environmental exposure to nephrotoxins. The main focus of this research is to examine the potential relationship between environmental exposure to known nephrotoxins including arsenic, cadmium and lead and the potential health risk associated with the progressive dysfunction of the kidneys in renal impaired patients with CKD across Northern Ireland. In addition to these known nephrotoxins, co-abundance with several essential elements has been found to play a role as protecting mechanisms while others increase the uptake of nephrotoxic elements as a result of similar absorption mechanisms within the body. Key elements protecting the body from toxicity include selenium and zinc, whereas those which have been attributed to enhance the uptake of arsenic, cadmium and lead include iron and calcium. The compositional nature of the soil and stream geochemical data is explored to aid in the analysis of interactions between elements. Two approaches, one data-driven and the other knowledge-driven, are explored to investigate the associations between co-abundant elements. The bioaccessibility of these elements, which is the portion of the relevant toxin absorbed within the

1

body, is also investigated to identify areas across Northern Ireland with an increased environmental hazard and potential health risk. The study uses a combination of datasets from the United Kingdom Renal Registry (UKRR) unknown aetiology subset, the soil and stream geochemical dataset from the Tellus Survey (GSNI) with the addition of a bioaccessibility subset. Findings suggest a relationship between the presence of elevated arsenic in stream waters and impaired renal function of the kidneys. Interactions between essential elements and potentially toxic elements could explain the regional variation of CKD of uncertain aetiology across Northern Ireland.

Keywords

Environmental Toxins, Arsenic, Cadmium, Geochemistry, Bioaccessibility, Kidney Disease Data

1. Introduction

1.1 Environmental Toxins

The increasing global prevalence of Chronic Kidney Disease (CKD) has attracted attention with research focusing on the unknown aetiologies (causes) of the progression of kidney dysfunction. Patients presenting with CKD of "unknown aetiology" have developed CKD as a result of factors other than the commonly recognised factors of age, gender, ethnicity or pre-existing medical conditions. The unknown causes of renal disease are significant as treating every person with a variety of symptoms and diseases places a large burden on the National Health Service (NHS) budget within the UK. Insight into the spatial variation of CKD in relation to exposure to Potentially Toxic Elements (PTEs) may impact management and policy within the NHS and environmental services. The significant treatment cost of the disease places a great strain on the NHS and management of renal units within the United Kingdom (UK). In the UK, the prevalence of stages three to five of CKD is 8.5% with an average cost of renal replacement therapy, including dialysis and kidney transplants, of between £20,000 to £30,000 per patient per year comprising 1-2% of the

annual health budget (NHS, 2010; Lewis, 2012). Reviews of exposure to environmental toxins (Brewster, 2007 ; Soderland et al. 2010) have examined the multifactorial causes of CKD identifying several key nephrotoxins, including arsenic (As), cadmium (Cd) and lead (Pb), which affect the human body. The potential role of environmental toxins in the progression of CKD varies according to each element's mobility and solubility within the environment, exposure to the population and the nephrotoxic implications which include acute and chronic renal failure. The most toxic forms of these elements, which are involved in the progression of kidney disease, are often inorganic and become exposed to the population through environmental exposure by direct ingestion. The solubility and mobility of each trace element alters the bioaccessibility which is defined as the fraction that is soluble in the gastrointestinal environment and is available for absorption (Wragg and Cave, 2002). The co-abundance of toxic elements and essential elements may be associated with the progression of CKD. Some essential elements play a role as *protecting* mechanisms against toxic elements while others *increase* the uptake of toxic elements as a result of similar absorption mechanisms. The key elements which protect the body's cells from toxicity include zinc (Zn) and selenium (Se) and those which enhance uptake of As, Cd and Pb include iron (FeO) and calcium (CaO). The interactions between each of these elements provide insights into the effects of co-abundance and the progression of CKD. Mobility, solubility, bioaccessibility and co-abundance of these elements are involved in the potential progression of kidney dysfunction.

1.2 Exposure pathways

Over the past years, a risk-based approach with regard to the assessment of potential exposure scenarios to identify elevated soil borne PTE concentrations within the context of contaminated land assessments has been established in a number of countries (Rothstein et al., 2006). Within the UK for example, this risk-based assessment is completed in accordance with guidance documents developed by the Environment Agency (EA) using the Contaminated Land Exposure Assessment (CLEA) model (Jeffries and Martin, 2009). Common to both the wider investigation of human

3

health and welfare effects associated with naturally occurring variations in PTE concentrations and the specific assessment of anthropogenic contaminated land exposure scenarios to human health, is the underlying assessment of Source–Pathway–Receptor linkages. The above CLEA model identifies a number of relevant exposure pathways, including:

- Inhalation of outdoor and indoor dust and vapour;

- Dermal contact with soil and dust;

- Consumption of contaminated home grown produce and soil adhering to produce, and

- Ingestion of soil and dust.

While acknowledging that a number of the above exposure pathways may be present for individual PTEs across different exposure scenarios, this ecological exploratory study investigates the potential linkages between observed total PTE concentrations in soil and renal disease, focusing on the soil ingestion exposure pathway. Ingestion of soil can occur due to either intentional (geophagy or pica behaviour) or unintentional consumption by hand-to-mouth contact, dust ingestion, or from poorly washed home grown produce. Hand-to-mouth unintentional consumption can occur via adherence of soil particles to hands; the <250 μm soil fraction is considered the upper particle size fraction for which this might occur (Choate et al. 2006, Yamamoto et al. 2006, Sicialiano et al. 2009). The UK Environment Agency recognizes direct and indirect soil and dust ingestion and consumption of home grown produce as one of the key exposure pathways for residential land uses (Jeffries and Martin, 2009). To better assess exposure via the oral ingestion pathway, in vitro methods have been developed to estimate the oral bioaccessibility of soil contaminants. With regard to oral intake as a significant PTE exposure pathway in environmental exposure scenarios, this bioaccessible fraction refers to the PTE fraction released in the gastro-intestinal (GI) tract by digestive juices, thus representing the maximum contaminant concentration available for intestinal absorption (Cave et al. 2011). Other environmental exposure pathways for individual PTEs, such as dermal contact with

4

and ingestion of water-borne PTEs (eg. in private drinking water supplies) may furthermore affect health outcomes with regard to renal disease. This exposure pathway is however beyond the scope of this exploratory study, focusing on the ingestion exposure pathway of soil-borne PTEs.

1.3 Rationale and Approach

Previous studies have concentrated on the relationships between potentially toxic elements (PTEs) and a range of diseases (Ferreccio and Sancha, 2006; Meliker et al. 2007; Navas-Acien et al. 2009; Ryu et al. 2010; Horton et al. 2013). One such study based in Northern Ireland studied the bioaccessibility of trace elements in the soil and the relationship between several cancers (Barsby et al. 2012). This study involved assessing the bioaccessibility of elements which showed that certain areas with elevated total concentrations of elements (elevated nickel in soils over Palaeogene flood basalt areas) had low bioaccessibility presenting lower risk to the population and that areas of moderate total elemental abundance showed high bioaccessibility (moderate levels of arsenic in metasediments). The focus of this paper is to investigate potential associations between environmental toxins and the progression of CKD of unknown aetiology, in an attempt to explain the regional variation of prevalence rates of CKD across Northern Ireland as provided by the UK Renal Registry (UKRR).

This exploratory paper investigates the potential linkages between observed total PTE concentrations in soil and renal disease, focusing on the soil ingestion exposure pathway, including intentional and unintentional consumption of soil. To this end, the study explores the 'rural' soil geochemical dataset provided by the Tellus Survey (Smyth 2007), covering the whole of Northern Ireland (approx. 13,850 km^2) with a sampling density of one sample site every 2 km^2 across the rural areas of Northern Ireland. Census data over the past decades highlight that the population across rural Northern Ireland exhibits a relatively low mobility rate in comparison with other regions of the UK. Champion (2005) noted that Northern Ireland recorded by far the lowest rate of movement across regional and country boundaries within the UK with 90.7% of residents remaining at the same

residence in the 12 months prior to the 2001 census. To this end, Northern Ireland provides an ideal region for studying potential linkages between the spatial distribution of elevated soil-borne PTE concentrations and health outcomes with regard to renal disease via the ingestion exposure pathway as described above. It is important to note that the presented study does not attempt to analyse the potential complex exposure pathways along the commercial food-chain. Such a study would require amongst others a comprehensive assessment of food-borne PTE concentrations as well as detailed dietary intake information, which is beyond the scope of this paper. While acknowledging that these exposure pathways may contribute to health outcomes with regard to renal disease, the presented study is exploring potential linkages between observed elevated total and PTE concentrations in soil and renal disease, focusing on the oral ingestion exposure pathway for residential land use. The presented study is aimed at exploring potential linkages between soil-borne total and bioaccessible PTE concentrations and renal disease in order to address the observed rates of unknown aetiologies. To this end, the study aims at demonstrating the potential benefits for using existing comprehensive environmental datasets to provide additional information which may help to improve health outcomes and serve as an exploratory basis for further more detailed studies into potential environmental factors for renal disease.

1.4 Compositional Data Analysis

This research explores the potential relationship between renal disease patterns and the abundance of PTEs in soils, stream sediments and stream waters. However, the geochemical components for soil, stream sediments and stream waters, form a multivariate system that should be analysed as a whole, not component-wise (McKinley et al. 2016). Namely, geochemical data are observations that contain quantitatively expressed relative contributions of parts on a whole. The relevant information is conveyed by the ratios between the components; it does not matter if the components add up or not to a constant value. Therefore, even with a variable sum of geochemical concentrations, the relative nature of geochemical data is still present and needs to be taken into account (McKinley et al. 2016).

Therefore, a compositional data approach was explored to comply with the relative nature of the geochemical data (Aitchison, 1986; Pawlowsky-Glahn and Buccianti, 2011; Pawlowsky-Glahn, Egozcue, Tolosana-Delgado, 2015; van den Boogaart and Tolosana-Delgado (2013); Templ et al., 2011). McKinley et al. (2016) caution that if a compositional approach is not used, it becomes difficult to guarantee that any identified pattern, trend or anomaly is not an artefact of the compositional nature of the data. This is particularly important for exploring the role of environmental factors on health where it is essential to understand how the variation in geochemical concentrations of PTEs may be influenced by dilution, enrichment or depletion in other elements. Exploratory data analysis for multi-element geochemical datasets is most often undertaken to explore patterns, links and associations between elements as part of the 'Discovery Process' (Grunsky et al., 2014). This paper adopts an initial, data-driven approach to capture patterns and processes that may be hidden or masked by dominant processes. A knowledge driven log-ratio approach is also explored to examine soil, stream sediments and stream water geochemistry data for a range of primary known nephrotoxins, including As, Pb and Cd, and secondary elements including Fe, Ca, Se, and Zn. Elemental associations and subcompostions of components from soils, stream sediments and stream waters geochemistry are identified and are used as environmental covariates for regressing CKD of unknown aetiology. Poisson regression analysis was used to explore the relationship between the environmental and disease data as the primary variable of interest, incidence rates of CKD measures abundance.

2. Study Site and Materials

Several factors make Northern Ireland an ideal region for research within the field of medical geology. The primary factor is Northern Ireland's complex and diverse geology which provides a representation of typical parent rock types and soils (Fig. 1a) and which has been mapped by the Geological Survey of Northern Ireland (GSNI) in the form of an available soil geochemistry dataset

known as the Tellus Survey (generated during 2004-2006). The diverse geology increases the transferability of the study to other places. A combination of the UKRR, containing data collected from 2006-2014 on end stage renal disease patients offers the potential to investigate the health risk of CKD associated with environmental toxins from the GSNI Tellus Geochemical Survey.

2.1 Disease Data

CKD is increasing annually in prevalence within Northern Ireland with an average prevalence of 3% in 2007 rising to 4.8% in 2014 and an age standardised incidence rate of 25.54 people per 100,000 (NISRA, 2014; data from UKRR, 2014). While an annual increase in CKD has been linked to increasing diabetes rates and longer life expectancy, the rates of unknown aetiology remain of concern. The unknown aetiology in renal disease patients in Northern Ireland was 17.2% of the total incidence rates in 2010; from 2006-2014 the age standardised incidence rate was 4.13 per 100,000 persons (Gilg et al. 2012; data from UKRR, 2014). This research uses a subset of data from the UKRR for 259 patients presenting with CKD of unknown aetiology with end stage renal disease indicating the requirement for dialysis and renal replacement therapy (Fig 1b and 1c). The data were combined with 890 super output areas (SOAs), which are the smallest census record district for Northern Ireland containing population counts based on the 2011 census data (NISRA, 2011). Cases recorded as unknown aetiology represent patients presenting with CKD, but where the medical practitioner deemed the cause unknown and not easily related to other known causes of CKD. The observed cases of CKD of unknown aetiology therefore provide the basis for an analysis between the spatial variation of CKD incidence and environmental trace element concentrations.

2.2 Geochemistry

The Tellus Survey was initiated in 2004 by GSNI to provide fine spatial resolution geochemical datasets (Smyth, 2007). Tellus Survey data types collected included stream sediment, heavy mineral

8

concentrate, stream water, surface soil and deep soil samples. These samples were analysed using several types of spectrometry methods to separate and identify the components of major and trace elements within every sample. Methods of analysis included X-ray fluorescence (XRF) spectrometry, Inductively Coupled Plasma (ICP) and ion-chromatography. A total of 5,874 sites were sampled for stream sediments and waters giving a distribution of one site per 2.4 km^2 (Smyth, 2007). The stream water samples were collected upstream of the stream sediments to avoid contamination (Young and Donald, 2013). Two types of soil samples were collected at each site including shallow surface soils from the 'A' horizon, which are used in this research, and deep soils from the 'B' horizon. Each sample was collected using a soil testing kit which included a hand auger to measure a 20 by 5 cm flight from a standard depth of 5-20 cm for surface soils and 35-50 cm for deep soils. In rural areas, 6,862 sites were sampled at an average of one site per 2 km^2 and in urban areas including Belfast 1,315 sites were sampled at a distribution of four sites per km^2 (Smyth, 2007; Young and Donald, 2013). The rural soil sediment data, as measured by XRF, along with the stream sediments and stream water datasets were used in this study. Published detection limits were used to replace zeros and negative amounts, however elements with a large number of censored values were excluded from further analysis. The full soil composition contained 47 elements; 8 elements were excluded including Sc, Mo, Nd, Bi, W, Ta, Ga, and SO3 due to censored values. The non-censored data sets used for the stream waters and sediments contained 16 elements. Details of the number of samples and summary statistics for the geochemical variables of As, Cd, Pb, Fe2O3, CaO, Se and Zn for each of the datasets (soil, stream sediment and stream water) are provided in Table 1.

2.3 Bioaccessibility

A subset of surface soil samples from the rural Tellus data was used for testing bioaccessibility, which is the proportion of the toxin solubilised (Wragg and Cave, 2002; Hrudey et al. 1996). Bioaccessibility data testing was undertaken for 145 of the previously collected and stored Tellus soil samples. The locations, shown in Fig. 1A, were selected to cover the range of rock and soil types

represented across Northern Ireland (Barsby et al. 2012; Palmer, et al. 2013). The method used for assessing the bioaccessibility of soil samples was the Unified BARGE Method (Wragg et al 2009; BARGE – INERIS, 2011). The method represented the oral exposure pathway through the stomach, and stomach with intestine using synthetic materials in place of the digestive system including gastric fluid, saliva, duodenal fluid and bile (Barsby, et al. 2012). This research used the bioaccessible fraction (BAF %) analysed by Barsby et al. (2012) and Palmer, et al. (2013) which was calculated from the Unified BARGE Method test for each sample using the highest bioaccessible PTE concentration of the digestive extracts. Full details of the Unified BARGE Method used to produce the BAF data used in this study are provided in Barsby et al. (2012). Bioaccessibility provides an indication of associated environmental hazard from the PTE to the population. This provided bioaccessible fraction data (BAF%) for the nephrotoxins of As, Cd, and Pb for 145 sample locations across North Ireland (Fig. 1A). Due to the compositional nature of the bioaccessibility fraction data (BAF%), a compositional data analysis approach was used for these data.

3. Methodology

3.1 Analysis of Geochemical Data

Summary statistics from exploratory data analysis of the primary known nephrotoxins, As, Pb and Cd, and secondary elements Fe, Ca and Zn for total soil, stream sediments and stream water concentrations are shown in Table 1. Dot maps representing the "measured" raw or absolute geochemical values were used to provide an initial visualisation of the geographic variation in high and low total concentrations and the bioaccessible fraction of the known nephrotoxins As, Pb and Cd across Northern Ireland (Fig 2). The "measured" raw dot maps were produced with the understanding and acknowledgement of dependence to all other variables and were used as an initial visualisation. Previous work by Barsby et al. (2012) showed that elevated total PTEs concentrations did not necessarily reflect high bioaccessibility, whereas areas of moderate total elemental

abundance of some PTEs such as As showed high bioaccessibility, Therefore, the BAF% data were compared with the dot map for PTEs (Figure 2) to assess the risk to human health in areas where clusters of elevated PTEs concentrations were present to investigate the regional variation of prevalence rates of CKD across Northern Ireland.

3.2 Compositional Data Analysis

The two approaches used for compositional data analysis included a data-driven approach and a knowledge-driven approach.

The centred log-ratio transformation (clr) was used to produce a compositional biplot and screeplot as part of the data exploration discovery process. The compositional biplot is obtained as a standard covariance biplot for the centered log-ratio (clr) data (Pawlowsky and Egozcue 2011). In the covariance biplot, up to a good projection, the length of a link between arrows is proportional to the standard deviation of the corresponding simple log-ratio of the elements of the corresponding arrows. Therefore, it is important to note that the compositional covariance biplot cannot be interpreted in the same way as a non-compositional principal component analysis (PCA).

Compositional Q-mode cluster analysis was used to investigate clustering of variables. The Q-mode hierarchical cluster analysis of chemical elements provide an orthonormal basis. The ilr-coordinates with respect to this basis try to approximate the scores of compositional principal components in a more interpretable way (Egozcue et al., 2003; Pawlowsky-Glahn et al. 2011a; Pawlowsky-Glahn et al. 2011b). These ilr-coordinates are called principal balances. The hierarchical cluster dendrograms shown in this paper were produced with a Ward cluster analysis with the variation matrix as distance between elements. The procedure is detailed in van den Boogaart and Tolosana-Delgado (2013).

3.2.1 Data-driven approach

Even in the absence of prior or expert knowledge, multivariate analysis techniques can help to identify interesting log-ratios and subcompositions to explore for further analysis. The advantage of a

11

data-driven compositional approach is that patterns and processes can be revealed that may be hidden or masked by dominant processes. The data-driven approach used the full geochemical composition to analyse and identify further relations between elements using compositional PCA and Compositional Q-mode cluster analysis. For the data-driven approach, balances in the hierarchical cluster dendrogram should highlight processes differentiating between similar elements. These may represent second-order processes that would otherwise be overshadowed by the major processes (McKinley et al. 2016).

3.2.2 Knowledge-driven approach

Single component elemental maps including the "measured" raw dot maps, although used widely, may not be sufficient to investigate the richness of the compositional geochemistry data. In particular for this study, the literature on renal disease indicated elemental associations between primary PTEs and secondary elements that cannot be determined fully by single component elemental mapping. It may, therefore, be more useful to ratio the PTEs to some other compositional entity. Therefore a knowledge driven log-ratio approach was explored to examine the soil, stream sediments and stream water geochemistry data for the primary known nephrotoxins, As, Pb and Cd, and secondary elements Fe, Ca, Zn and Se. These key nephrotoxins and secondary elements were used to form a knowledge driven sub composition of As, Pb, Cd, Fe, Ca, Zn and Se.

3.3 Regression

Poisson regression, which is appropriate for modelling count data (Cameron and Trivedi, 1998), was used to investigate any potential relationship between observed cases of CKD with unknown aetiology and environmental covariates (specifically soils, stream waters and stream sediments). The count of observed cases of unknown aetiology y_i defines the response variable, while the log-ratios x_i are the covariates in the regression. The Poisson regression model is then defined as:

12

$y_i \sim$ Poisson (λ_i)

with, $\ln(\lambda_i) = \beta_0 + \beta_1 x_1 + \ldots + \beta_i x_i$ for parameter λ (1)

where, the β are the linear fixed effect coefficients to be estimated.

Count variables are often associated with an exposure variable, which indicates the upper limit on the number of times the event could have happened (Cameron and Trivedi, 1998). Population was, thus, incorporated into the Poisson regression model as an offset as is standard practice. The glm function in R (R Development Core Team, 2011), which fits the Poisson model by maximum likelihood, was used in this study as follows:

glm(formula = obs ~ offset(log(Pop)) + ilr(data), family = poisson, data)   (2)

The Poisson regression model was, thus, fitted to the observed cases of unknown aetiology with the population as an offset and the Tellus geochemical concentrations as the covariates, specifically soils, stream waters and stream sediments.

The observed cases of CKD with unknown aetiology were provided for each Super Output Area (SOA; totalling 495 SOAs). The mean geochemistry value for each geochemical element was calculated per SOA. Four approaches were investigated. Initially the globalised linear model was applied to the mean geochemical data per SOA without applying log-ratios to the geochemical data to provide a benchmark against which to compare the log-ratio results. Following this non-compositional approach, Poisson regression analysis was applied based on information provided by the data driven and knowledge driven approaches. For the data driven approach a subcomposition was selected on the basis of those elements identified as significant subcompositions from the clr

13

covariance biplot and ilr dendrogram. For the knowledge driven approach, the subcomposition of As, Pb, Cd, Fe, Ca, Zn and Se was used for Poisson regression. The explanatory power of the resultant model is measured by its associated *p* value and calculated confidence interval (van den Boogaart and Tolosana-Delgado 2013).

4. Geochemical Results and Discussion

4.1 Spatial Distribution of Elemental Concentrations

The spatial distribution of each PTE varied across Northern Ireland exhibiting some relationship with the underlying geology (Fig. 1A and Fig. 2). Arsenic concentrations were low across flood basalts, the Antrim Lava Group and elevated for lithic sandstones and shales (Fig. 1 and Fig. 2A). High concentrations of As were present in the south and west of Northern Ireland. This pattern can be observed also in the stream sediment data (Table 1). Elevated soil concentrations of As across the south western region of Northern Ireland show anthropogenic contamination from agricultural lime (Young and Donald, 2013). The highest PTE As concentrations were located across the Dalaradian Supergroup of County Tyrone (Central Highland Grampian Terrain Fig. 1A) which had low to moderate bioaccessibility (Fig. 2D; Mitchell, 2004). Arsenic had the lowest BAF range (5-58%) of the three PTE concentrations (Fig. 2D, 2E and 2F). Previous studies (Barsby et al 2012; McKinley et al 2013) found that highest BAF levels of As were associated with the Gala Group metasediments in the Southern Uplands-Down Longford Terrain and these areas indicated a potential relationship between stomach cancer and the environmental impact of elevated As concentrations and BAF (Fig. 1A and Fig. 2D).

Elevated soil PTE concentrations of Cd were present within south Armagh and Fermanagh; these highlighted areas are also observed in the stream sediments and stream water data (Fig. 2B). The stream sediments indicate further elevated areas of Cd in the west of county Tyrone and north east of county Antrim although these areas have a low BAF (Fig. 2E). Cadmium has the largest range and

highest BAF (14-95%) with significant areas of elevated Cd soil concentrations located within regions with a low-to-moderate BAF (Fig. 2B and Fig.2E).

The distribution of Pb across the eastern region of Northern Ireland shows elevated concentrations within the soils across the Antrim Basalts, Palaeogene intrusive centres of the Mourne Mountains, the Newry Igneous Complex and the central highlands of the Sperrin Mountains (Fig. 1A; Mitchell, 2004). The natural distribution of Pb levels were associated with mineralisation in areas of moorlands and regions with historical contamination such as south Armagh mining district (Fig. 2C, Young and Donald, 2013). The key areas of elevated concentrations had a low to moderate BAF (8-68% BAF; Fig. 2F). Moderate to high BAFs in soils were also present across several of the elevated areas of lead within the stream water and stream sediments. The main elevated concentrations of Pb in the stream sediments were located in south Armagh and Down with a few hotspots in north Antrim and West Tyrone.

4.2 Data driven approach

For the soil and stream sediments data the proportion of variance explained by the first two components in the compositional covariance biplot (Fig. 3A and Fig. 4A) and screeplot (Fig. 3B and Fig. 4B) is 64.65% and 60.45% respectively. In the covariance biplots the length of the rays is approximately proportional to the variance of the clr-components. For the soil data, the clr of elements including MnO, Ni, Co, Fe2O3, Cr point to the right (positive side of the first principal component) while the key nephrotoxins of As, Pb and Cd, amongst other elements including U, Se and K2O, point to the left (Fig. 3A). This is reversed for the stream sediments data (Fig. 4A). In both cases, the second component may discriminate nephrotoxins quite well from the other elements. Subcompositions are indicated by short links (the two proportional variables have quasi-constant log ratios, the links are short and the arrow heads lie together (van den Boogaart and Tolosana-Delgado 2013). Several subcompostions can be observed from the soil data covariance biplot. The subcomposition (MgO, Zn, CaO) has a metric variance of 0.09 (1.45% of the total variance); the

15

subcomposition (Na2O, SiO2, Al2O3, K2O) has a metric variance of 0.12 (1.94% of the total variance); the subcomposition (MnO, Ni, Co, Fe2O3, Cr) has a metric variance of 0.23 (3.72% of the total variance) and the subcomposition (As, Pb, Cd) has a metric variance of 0.19 (3.07% of the total variance) (Pawlowsky-Glahn and Egozcue, 2001). The observed subcompositions for the soil data are interesting in that the known nephrotoxins form a subcomposition (As, Pb, Cd with other trace elements and Se). The other subcompositions include some of the elements Ca, Zn important in their role as inhibitors or where their co-abundance with the nephrotxins has been related to the potential progression of kidney dysfunction.

The soil data cluster dendrogram provides further information on balances and potential subcompositions (Fig 3C). Some of the balances include specific elements which interact within the body as identified from a broad literature review. The cluster dendrogram for the stream sediment data (Fig 4C) shows a large balance which includes secondary essential elements which exhibit defensive properties within the body including Zn and CaO against two of the key known nephrotoxins As and Pb (Alloway, 1995; Garcia-Rico, et al. 2012; Tchounwou, et.al 2003).

4.3 Knowledge driven approach

Using a knowledge driven approach, the scree plot for the Tellus soils shows that the first two components explain 80% of the total variance. The first subcomposition (Zn, Feo, CaO) groups the essential elements and the second subcomposition (As, Cd and Pb) with a total variance of 35.62%. This concurs with results using the data driven approach where the subcomposition As, Cd and Pb was identified.

5. Poisson Regression

Poisson regression analysis was carried out using a generalised linear model for initial data analysis before the addition of the ilr transformation. Using compositional data analysis the ilr transformation

within the Poisson regression model was used to determine the relations and significance between the elements and observed cases of CKD.

5.1 Non-compositional Approach

A non-compositional approach was used for initial analysis of the results before an ilr transformation was applied. The significance for each Poisson regression model increased for various elements including Ca and Zn when a compositional approach was applied compared with the non-compositional approach (Tables 2 and 3). Only the essential elements FeO, Zn and Se, where p-values were less than a significance level α=0.05, were identified as significant using a non-compositional approach (Table 2). Compositional analysis of the variables (using ilr transformation) allowed the results to remain unaffected by the compositional nature of the geochemical data and the results demonstrate a larger number of variables where p-values were less than a significance level α=0.05 (Table 3).

5.2 Knowledge Driven Approach

The knowledge driven approach was based on known PTEs which impair the kidneys. These key PTEs included Cd, As, and Pb. The geochemical variables when analysed using the knowledge driven approach with the nephrotoxic subcompositions indicated the highest number of variables where p-values were less than a significance level α=0.05 between the observed cases of unknown aetiology in CKD with log ratios of MgO, Zn and CaO mean concentrations. The regression model highlighted the log ratio Zn, over FeO, CaO, Pb, As and Cd, as the most significant element associated with the progression of CKD. Zinc has been found to increase the uptake of cadmium in plants, forming increased risk of exposure in soils with elevated zinc concentrations. Ingestion of food with accumulated concentrations of cadmium is a risk factor for the progression of CKD in Sri Lanka where the incidence rate of CKD of unknown aetiology is increasing (Jayatilake , et al., 2013;

Jayalal, 2015). Within the subcomposition the interactions between elements are significant as these effect the overall environmental exposure risk to PTEs.

The stream sediments indicate elevated areas of both As and Cd in County Tyrone and County Armagh but elsewhere As and Cd show different spatial distributions (Fig. 2). The role of iron may be important in understanding the link between As and Cd and the strength of relationship with observed CKD cases of unknown aetiology. The elemental mobility of As in the environment is normally low (Young and McDonald 2013) but As is known to become soluble and mobile under reducing conditions with iron oxides (ATSDR, 2007). Arsenic is strongly absorbed onto the surface of hydrous iron oxides in a low pH environments (Odor et al. 1998). Young and McDonald (2013) state that in the absence of Fe and Mn oxides in stream sediments, such as over limestone bedrock, As may remain in solution and have greater mobility. Cadmium, on the other hand is very soluble and mobile at low pH values. Excess Mg and Ca added to stream waters may increase the soluble Cd content but inhibit the release of Fe due to competition for absorption sites on clays by cations (Curtis and Walker 1994). Moreover, it has been observed that Fe deficiency in the body causes an up regulation of the Fe channels within the body through which both Cd and Fe are absorbed into the digestive tract (Thomas, et al., 2009). It is clear that this finding requires more work to explore these complex relations further. The effects of these elements within water supply and the spatial variation of CKD have been the focus of study in global areas including Nicaragua and Sri Lanka (Jayasumana, et al., 2014; Jayatilake et al., 2013). A factor, therefore, to consider is the degree of access to the mains water supply compared with domestic well supplies across Northern Ireland. Private groundwater supplies account for only a small portion of the population (<1% with 122 registered and an estimated 4,000 unregistered private supplies).

The relations between these elements are relevant within the study of environmental exposure as a deficiency of an essential element can cause the body to uptake PTEs such as As, Pb and Cd. Fe and

Ca are significant coexisting elements and people who suffer from iron deficient anaemia and calcium deficiency are at risk of being exposed to several potentially toxic elements including cadmium, arsenic and lead as they enter the body through similar mechanisms (Turgut et al, 2009; Martin, De Burca, & Morgan, 2009; Selinus, et al., 2013; Jarup, 2003).

In addition to the presence of a significance level less than $\alpha=0.05$ for the log ratio of Zn, Ca was also found to be significant in stream waters (Table 2 and Table 3). As these cases of CKD are associated with end stage renal disease patients, the kidneys have reduced function and often an excess of essential elements can lead to kidney problems such as kidneys stones; Ca is often found as contributor to problems with kidney stones (Selinus, et al., 2013). These results could indicate an excess of Ca and Zn may contribute to the observed cases of CKD of unknown aetiology.

5.3 Data Driven Approach

The results obtained from the data driven approach showed less relations between the potential progressions of CKD of unknown aetiology with exposure to environmental elements. The key result using the data driven approach was found in the analysis of the full geochemical soil composition for MgO (magnesium oxide) (Table 3) where p-values were less than a significance level $\alpha=0.05$. Larger subcompositions identified from the compositional covariance biplot and cluster dendrogram indicated no significance within the UKRR subset. These results indicated the knowledge driven sub compositional approach was more useful in determining the significance of the relationship between CKD and environmental toxins.

6. Conclusion

The Poisson regression model was fitted to the relations between the prevalence of CKD of uncertain aetiology and PTEs including arsenic, cadmium and lead in the soils, stream sediments and stream waters of Northern Ireland. Following data analysis of the Tellus geochemical data in Northern

Ireland the relations between CKD of unknown aetiology and environmental covariates were examined. Previous research showed that the progression of CKD can occur through exposure to environmental toxins in moderate and low concentrations. The most toxic forms of these elements involved in the progression of kidney disease are often inorganic and the population becomes exposed to them through direct ingestion of water or food containing the element and inhalation from occupational hazards. The key result indicated a relation between CKD of unknown aetiology and the environment. The main elements associated with CKD, where p-values were less than a significance level $\alpha=0.05$, included elevated concentrations essential elements zinc and calcium. The compositional approach to the analysis of CKD with PTEs was appropriate as it indicated variables of significance in relation to the composition or relevant subcompositions. Further influences from socio-economic, water sources and supply areas across Northern Ireland should be considered as potential factors contributing to the spatial variation of CKD. Further research of the co-abundance of PTEs, including examination of other socio-economic factors, the level of access to public water sources and the impact within the renal system will provide insight into the progression of CKD of unknown aetiology.

Acknowledgements

Figure Captions

Figure 1(A) Simplified geology adapted from McKinley et al. 2013, modified after Mitchell 2004; The sample sites indicate the Tellus soil samples which were analysed for bioaccessibility by Barsby et al. 2012 and Palmer et al. 2013; (B) Number of observed cases in CKD of unknown aetiology (UKRR, 2014); (C) Standardised Incidence Ratio of CKD patients with unknown aetiology.

Figure 2 Dot maps of the potentially toxic element soil concentrations (PTEs) in mg/kg as measured by XRF A) arsenic (As mg/kg); B) Cadmium (Cd mg/kg) and C) Lead (Pb mg/kg). Bioaccessibility Fraction BAF % for the PTEs of D) As (BAF%); E) Cd (BAF%); F); Pb (BAF%).

Figure 3 (A) Compositional covariance biplots; and (B) Cluster dendrogram for Tellus soil geochemistry data.

Figure 4 (A) Compositional covariance biplot; (B) Cluster Dendrogram for the Tellus stream water geochemistry data.

Table captions

Table 1

Table 1 Summary statistics for total concentrations of individual potentially toxic elements (PTEs) and protective essential elements in mg/kg as measured by X-Ray Fluorescent Spectrometry (XRF) for Tellus geochemistry sediment data, stream sediments and waters data.

Table 2 GLM model using non-transformed (raw) data for identified variables (As, Cd, Pb, Fe, Ca, Zn and Se) showing p values with a significance level less than $\alpha=0.05$.

Table 3 GLM using ilr transformation for the data driven approach for the full soil composition and relevant knowledge driven approach for the geochemical subcompositions with uncertain aetiology showing p values with a significance level less than $\alpha=0.05$.

References

ATSDR, 2007. Toxicological Profile for Arsenic, Atlanta: U.S. Department Of Health And Human Services.

Alloway, B. J., 1995. Cadmium. In: B. J. Alloway, ed. Heavy Metals in Soils. London: Blackie Academic & Professional.

Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press).

BARGE – INERIS, 2011 UBM Procedure for the Measurement of Inorganic Contaminant Bioaccessibility from Solid Matrices.

Barsby, A., McKinley, J., Ofterdinger, U., Young, M., Cave, M., Wragg, J., 2012. Bioaccessibilty of trace elements in Northern Ireland. Science of the Total Environment , 433, 397-417.

Brewster, U. C., 2007. Chronic Kidney Disease from Environmental and Occupational Toxins. Connecticut State Medical Society , 70 (4), 229-238.

Cave, M.R., Wragg, J., Denys, S., Jondreville, C., Feidt, C., 2011. Oral bioavailability. In: Swartjes FA, editor. Dealing with contaminated sites. Netherlands: Springer; 287–324.

Champion, T., 2005. Population Movement within the UK. In CHAPPELL, R. (ed.)

Focus on People and Migration. Palgrave Macmillan: Basingstoke, 91-113.

Choate, L.M., Ranville, R.F., Bunge, A.L., Macaladay, D.L., 2006. Integr. Environ. Assess. Manage. 2 (4), 375–384.

Curtis, P. J., Walker, P., 1994. Release of metals from a Cd contaminated streambed in Response to experimental Acidifcation and Netralization. Water Resources Research 30, 3449-3454.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. Mathematical Geology 35(3):279–300.

Ferreccio, C., Sancha, A. M., 2006. Arsenic Exposure and Its Impact on Health in Chile. The Journal of Health, Population and Nutrition, 24(2), pp. 164-165.

Garcia-Rico, L., Tejeda-Valenzuela, L., Velez, D., Montoro, R., 2012. Content of selenium, total and inorganic arsenic and bioaccessibility of arsenic in children diets of Mexico.. Science Food Agriculture, Volume 92, 1725-1731

Gilg, J., Castledine, C., Fogarty, D., 2012. UK Renal Registry 14th Annual Report: Chapter 1 UK RRT Incidence in 2010: National and Centre-Specific Analyses. Southhampton: Karger AG, Basel.

Grunsky, E.C., Mueller, U.A., Corrigan, D., 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: Applications for predicive geological mapping. Journal of Geochemical Exploration. 141, 15-41.

Horton, L. M., Mortensen, M.E., Iossifova, Y., Wald, M.M., Burgess, p., 2013. What Do We Know of Childhood Exposures to Metals (Arsenic, Cadmium, Lead and Mercury) in Emerging Market Countries?. International Journal of Pediatrics, 1-13.

Hrudey, S. E., Chen, W., Rousseaux, C. G., 1996. Bioavailability in Environmental Risk Assessment. New York: Lewis Publishers.

Jarup, L. 2003. Hazards of heavy metal contamination. British Medical Bulletin , 68, 167-182.

Jayalal, A. T., 2015. Chronic kidney disease of uncertain aetiology: adding vital piece of information to the national project team report of Sri Lanka. *BMC Nephrology* , 16(216).

Jayasumana, C., Gunatilake, S., Senanayake, P., 2014. Glyphosate, Hard Water and Nephrotoxic Metals: Are They the Culprits Behind the Epidemic of Chronic Kidney Disease of Unknown Etiology in Sri Lanka? International Journals of Environmental Research and Public Health , 11 (2), 2125-2147.

Jayatilake, N., Mendis, S., Maheepala, P., Metha, F. R., 2013. Chronic kidney disease of uncertain aetiology: prevalence and causative factors in a developing county. BMC Nephrology , 14 (180), 1-13.

Jeffries J, Martin I., 2009. Updated technical background to the CLEA model, SCHO0508BNQW-E-P.

Lewis, R., 2012. Understanding Chronic Kidney Disease: A guide for the non-specialist. England: M&K Update Ltd.

Martin, I., Morgan, H., Waterfall, E., 2009. Soil Guidance Values for Cadmium in Soil. Bristol: Environment Agency.

Meliker, J. R., Wahl, R. L., Cameron, L. L., Nriagu, J. O., 2007. Arsenic in drinking water and cerebrovascular disease, diabetes, mellitus, and kidney disease in Michigan: a standardized mortality ratio analysis. Environmental Health , 6 (4), 1-11.

McKinley, J.M., Tolosana Delgado, R., Hron, K., de Caritat, P., Grunsky, E., Reimann, C., Filzmoser P., and van den Boogaart K, G., 2016. Single Component map: Fact or Fiction? Journal of Geochemical Exploration, http://dx.doi.org/doi:10.1016/j.gexplo.2015.12.005

McKinley, J. M., Ofterdinger, U., Young, M., Barsby, A. & Gavin, A. 2013. 'Investigating local relationships between trace elements in soils and cancer data' Spatial Statistics, vol 5, pp. 25-41., http://dx.doi.org/10.1016/j.spasta.2013.05.001

Mitchell, I., 2004. The Geology of Northern Ireland Our Natural Foundation. Belfast: Geological Survey of Northern Ireland Belfast.

Navas-Acien, A., Tellez-Plaza, M., Guallar, E., Muntner, P., Silbergeld, E., Jaar, B.,

Weaver, V., 2009. Blood cadmium and lead and Chronic Kidney Disease in US Adults: A Joint Analysis. American Journal of Epidemology, 170(9), 1156-1164.

NHS, 2010. Kidney Disease Facts and Figures. United Kingdom: NHS.

NISRA, 2014. Disease Prevalence (Quality Outcomes Framework) (administrative geographies). http://www.nisra.gov.uk/ accessed 4th June 2014.

NISRA, 2011. Population Census SOA. http://www.nisra.gov.uk/ accessed 4th June 2014.

Odor, L., Wanty, R. B., Horvath, E. Fugedi, U., 1998. Mobilization and atentauton of metals downstream from a base-metal mining site in the Matra Mountains, northeastern Hungary. Journal of Geochemical Exploration, 65, 47-60.

Paleneeswaren, S., Abraham Sam Rajam, P. M., Silambana, S., 2013. Blood Arsenic and Cadmium Concentrations in End Stage Renal Disease Patients who were on Maintenance Haemodialysis. Journal of Clinical and Diagnostic Research, 7(5), pp. 809-813.

Palmer, S. Ofterdinger, U., McKinley, J.M., Cox, S., Barsby, A., 2013. Spatial Analysis Approaches to Investigate the Bioaccessibility of Nickel, Vanadium and Zinc in Northern Ireland, UK Soils. Environmental Geochemistry and Health, vol 35 (5), 569-584., http://dx.doi.org/10.1007/s10653-013-9540-0

Pawlowsky-Glahn, V., J. J. Egozcue, J.J., 2001. Geometric Approach to Statistical Analysis on the Simplex, Stochastic Environmental Research and Risk Assessment,15, 5, 384-398, 2001.

Pawlowsky-Glahn, V., Buccianti, A. (Eds.) 2011. Compositional Data Analysis, Theory and Applications. Wiley, Chichester (UK) 378p.

Pawlowsky-Glahn, V., Egozcue, J.J., 2011. Exploring Compositional Data with the CoDa-Dendrogram. Austrian Journal of Statistics, Volume 40, Number 1 & 2, 103-113

Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delegado, R., 2011a. Principal balances. Proceedings of the 4th International Workshop on Compositional Data Analysis, CoDaWork-2011, Egozcue, J. J., Tolosana-Delgado, R. and Ortego, M. I. (eds.) ISBN: 978-84-87867-76-7, Sant Feliu de Guixols, Girona, Spain.

Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delegado, R., 2011b. Principal balances to analyse the geochemistry of sediments.Proceedings of the 15th annual conference of the International Association for Mathematical Geosciences, IAMG-2011, Salzburg, Austria, www.iamg2011.at September 5-9.

Pawlowsky-Glahn, Egozcue, Tolosana-Delgado, Modelling and Analysis of Compositional Data, Wiley, 2015

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at http://www.R-project.org/.Rothstein H, Irving P, Walden T, Yearsley R., 2006. Environ Int; 32(8):1056–65.

Rothstein, H., Huber, M., Gaskell. G. 2006. A Theory of Risk Colonisation: The Spiralling Regulatory Logics of Societal and Institutional Risk, Economy and Society, 35, 1, 91–112.

Ryu, H., Chung, J. S., Nam, T., Moon, H. S., Nam, K., 2010. Incorporation of Heavy Metals Bioavailability into Risk Characterization. Clean-Soil, Air and Water, 38(9), 812-815.

Selinus, O., Alloway, B., Centeno, J. A., Finkelman, R. B., Fuge, R., Lindh, U., et al., 2013. Essentials of Medical Geology. London: Elsevier.

Siciliano, S.D., James, K., Zhang, G., Schafer, A.N., Peak, J.D., 2009. Adhesion and enrichment of metals on human hands from contaminated soil at an Arctic urban brownfield. Environ. Sci. Technol 43 (16), 6385–6390.

Smyth, D., 2007. Methods used in the Tellus Geochemical Mapping of Northern Ireland. OR/02/022

Soderland, P., Lovekar, S., Weiner, D. E., 2010. Chronic kidney disease associated with environmental toxins and exposures. Advanced Chronic Kidney Disease, 17(3), pp. 254-264.

Tchounwou, P. B., Ayensu, W. K., Ninashvili, N., Sutton, D., 2003. Environmental Exposure to Mercury and Its Toxicopathologic Implications for Public Health. Wiley Periodicals, Inc, 20 March, 149-175.

Templ, M., Hron K., Filzmoser, P., 2011. robCompositions: an R-package for robust statistical analysis of compositional data, In V. Pawlowsky-Glahn and A. Buccianti, eds. Compositional Data Analysis: Theory and Applications. Chichester, UK: John Wiley & Sons. Pp. 341-355.

Thomas, L. D., Hodgson, S., Nieuwenhuijsen, M. Jarup, L., 2009. Early Kidney Damage in a Population Exposed to Cadmium and Other Heavy Metals. Environmental Health Perspectives, 117(2), pp. 181-184.

Tolosana-Delgado, R., van den Boogaart, K.G., 2013. Joint consistent mapping of high-dimensional geochemical surveys, Mathematical Geosciences, 45: 983-1004.

Turgut, S., Hacıoğlu S., Emmungil, G., Turgut, G., Keskin, A., 2009. Relations between Iron Deficiency Anaemia and Serum Levels of Copper, Zinc, Cadmium and Lead. Polish J. of Environ. Stud., 18(2), pp. 273-277.

UKRR, 2014. Dataset of End Stage Renal Disease Patients from United Kingdom Renal Registry

van den Boogaart, K.G., Tolosana-Delgado, R., 2013. Analyzing compositional data with R. Springer, Heidelberg (Germany) 258p.

Wragg, J., Cave, M. R., 2002. In-vitro Methods for the Measurement of the Oral Bioaccessibility of Selected Metals and Metalloids in Soils: A Critical Review. Bristol: Environment Agency.

Wragg, J., M. R. Cave, H. Taylor, N. Basta, E. Brandon, S. Casteel, S., et al., 2009. Interlaboratory Trial of a Unified Bioaccessibility Procedure. OR/07/027.

Yamamoto, N., Takahashi, Y., Yoshinaga, J., Tanaka, A., Shibata, Y., 2006. Arch. Environ. Contam. Toxicol. 51, 157–163.

Young, M. E., Donald, A. E., 2013. A Guide to the Tellus Data. Belfast: Geological Survey of Northern Ireland, Belfast.

Table 1

Geochemistry		Min (mg/kg)	1st Quartile (mg/kg)	Median (mg/kg)	Mean (mg/kg)	3rd Quartile (mg/kg)	Max (mg/kg)	Standard Deviation (mg/kg)	Skew	Coefficient of variation
Soils **6,862 samples**	As	0.01	6.50	8.70	10.48	11.70	271.20	9.91	10.03	0.95
	Cd	0.20	0.40	0.50	0.56	0.70	63.30	0.81	68.53	1.45
	Pb	2.20	22.20	28.80	41.66	41	3707.50	234.23	75.06	5.62
	Fe$_2$O$_3$	0.30	2.79	4.19	4.65	5.69	42.25	2.85	1.50	0.62
	CaO	0.30	0.64	0.85	1.15	1.49	16.33	0.78	2.93	0.68
	Se	0.10	47.00	0.70	0.84	101.90	7.80	0.49	2.58	0.58
	Zn	2.80	0.60	71.80	78.35	1.00	2460.50	54.29	14.28	0.69
Stream **Sediments** **5,874 samples**	As	0.09	4.40	9.00	14.99	15.00	1400.00	34.72	18.87	2.32
	Cd	0.25	0.25	0.50	0.83	0.90	56.40	1.42	15.96	1.72
	Pb	0.70	21.90	30.00	35.53	39.00	1245.00	38.68	13.96	1.09
	CaO	0.34	1.11	1.87	2.91	3.85	38.43	2.91	3.74	1.00
	Fe$_2$O$_3$	0.73	5.19	6.76	7.61	9.92	50.08	3.39	1.62	0.45
	Se	0.01	0.20	0.40	0.56	0.70	30.20	0.71	15.99	0.79
	Zn	12	94.00	129.95	160.09	178.00	3161.50	147.59	8.08	0.92
Stream Waters **(mg/L)** **5,874 samples**	As	0.00005	0.0004	0.00079	0.00121	0.00138	0.04526	0.00179	0.0083	0.00148
	Cd	0.000001	0.000001	0.00001	0.00002	0.00002	0.00166	0.00005	0.0197	0.0025
	Pb	0.00005	0.00005	0.00008	0.00022	0.00021	0.04430	0.00076	0.0373	0.00346
	Fe	0.00001	0.00007	0.00022	0.00064	0.00064	0.04105	1.00072	0.0113	0.00269
	Ca	0.0002	0.01969	0.03502	0.04001	0.05228	0.276	0.02904	0.0013	0.00073
	Se	0.00001	0.00006	0.00018	0.00042	0.00059	0.00563	0.00058	0.0026	0.00138
	Zn	0.0002	0.00132	0.00230	0.00541	0.00463	0.363	0.01293	0.0106	0.00239

Table 2

GLM – Poisson Regression	GLM Coefficients			
	Variable	Intercept	Estimate	Pr(>|z|)
Knowledge Driven Subcomposition (Soils)	Se	0.9821	0.3313	0.00300
Knowledge Driven Subcomposition (Sediments)	Fe203	0.1349748	0.0512	0.00836
	Zn	0.0036938	0.0015	0.01468

Table 3

ILR Model- Poisson Regression	ILR Model Coefficients			
	Variables	Intercept	Estimate	Pr(>|z|)
Full Composition (Soils)	$\ln(MgO\char`\^18/(Na_2O*Ce*La*Ba*Cs*I*Te*Sb*Sn*In*Ag*Se*Zn*Fe_2O_3*CaO*Pb*As*Cd))$	1.8075	0.9139	0.0473
Knowledge Driven Subcomposition (Streams)	$\ln(CaO\char`\^3/(Pb*As*Cd))$	0.3757	0.1398	0.0072
Knowledge Driven Subcompositions (Sediments)	$\ln(Zn\char`\^5/(FeO*CaO*Pb*As*Cd))$	1.1985	0.4110	0.0035

A) As mg/kg
- 0.00 – 7.20
- 7.21 – 10.60
- 10.61 – 271.20

D) As BAF%
- 3.35 – 8.58
- 8.59 – 12.38
- 12.39 – 16.16
- 16.17 – 58.61

B) Cd mg/kg
- 0.20 – 0.40
- 0.41 – 0.60
- 0.61 – 63.30

Cd BAF%
- 0.04 – 15.59
- 15.60 – 33.40
- 33.41 – 52.60
- 52.60 – 95.57

C) Pb mg/kg
- 2.20 – 24.20
- 24.21 – 36.60
- 36.61 – 18759

F) Pb BAF%
- 0.17 – 15.61
- 15.62 – 23.12
- 23.13 – 36.00
- 36.01 – 68.63

A)

B)

A)

B)

Cluster Dendrogram

dd
hclust (*, "complete")
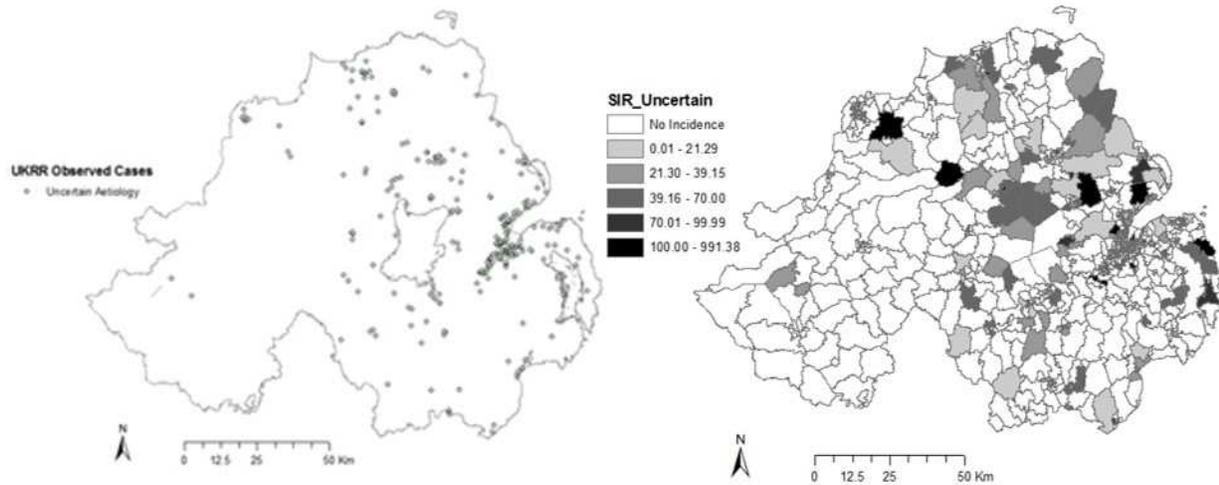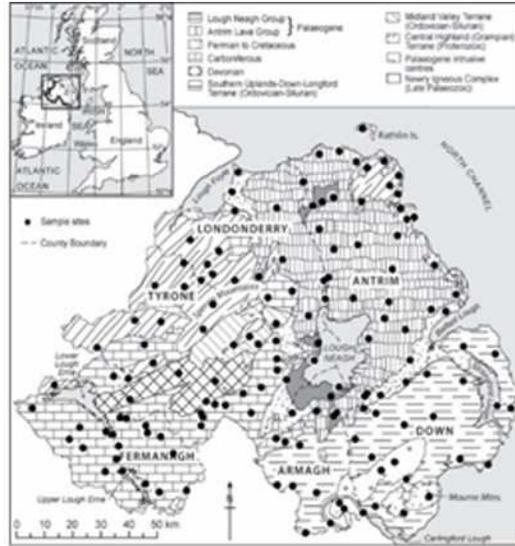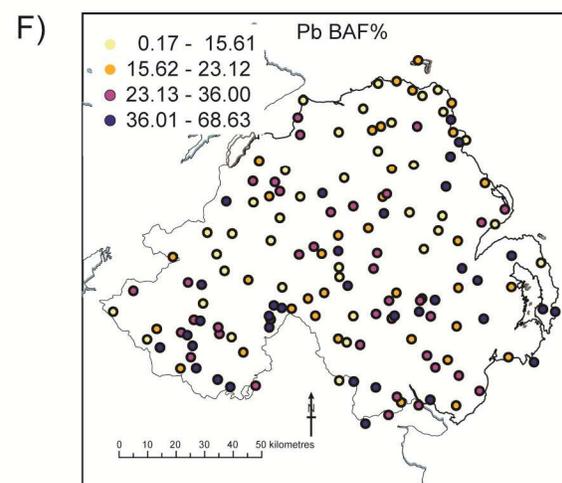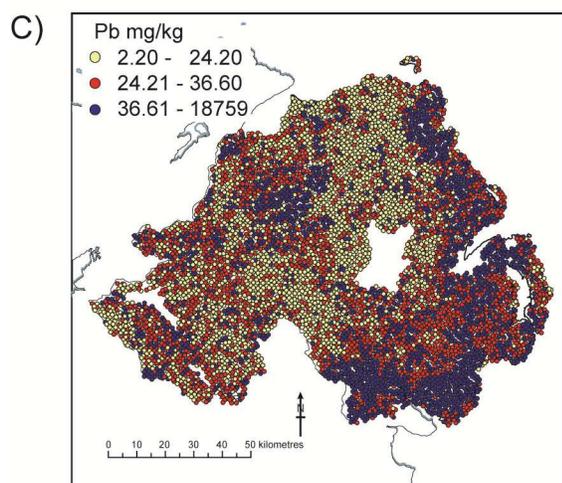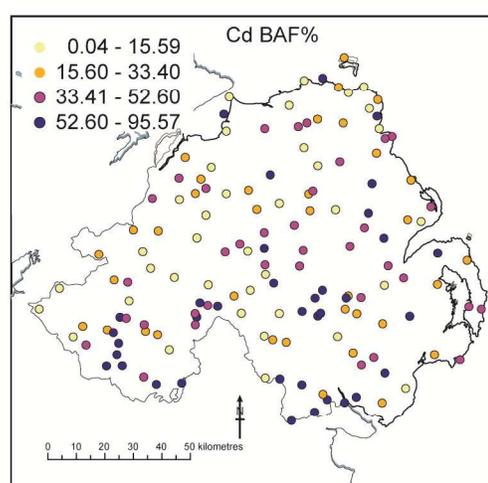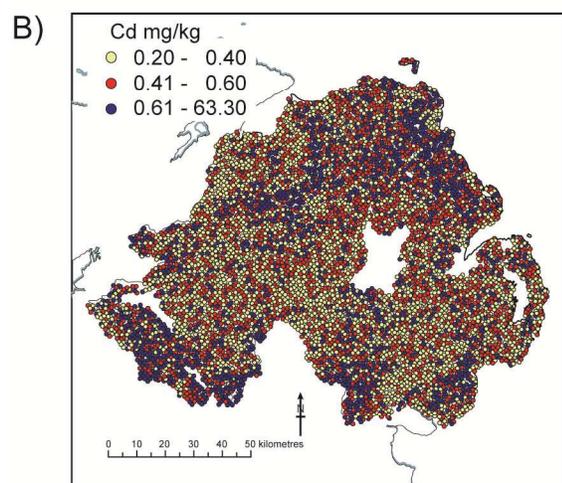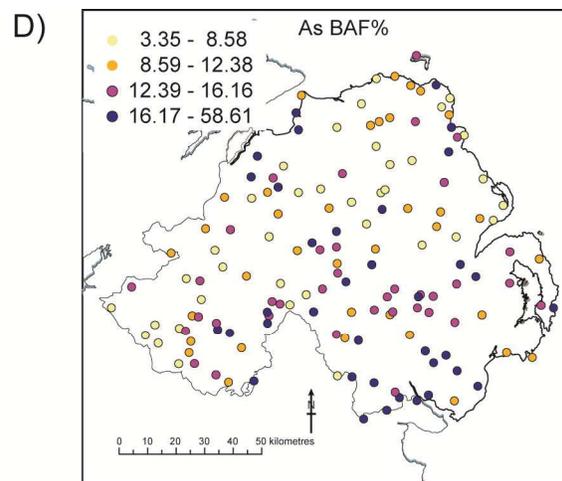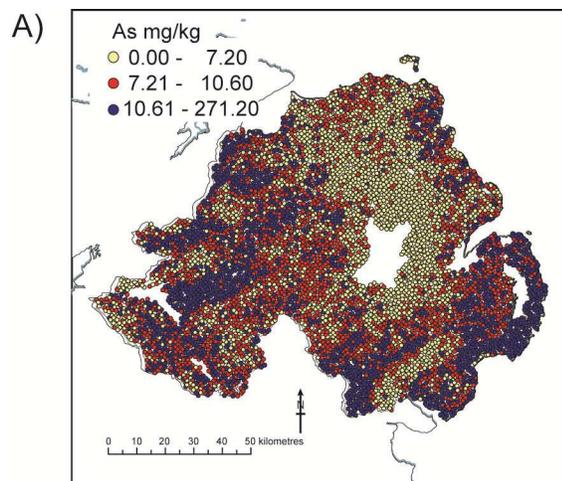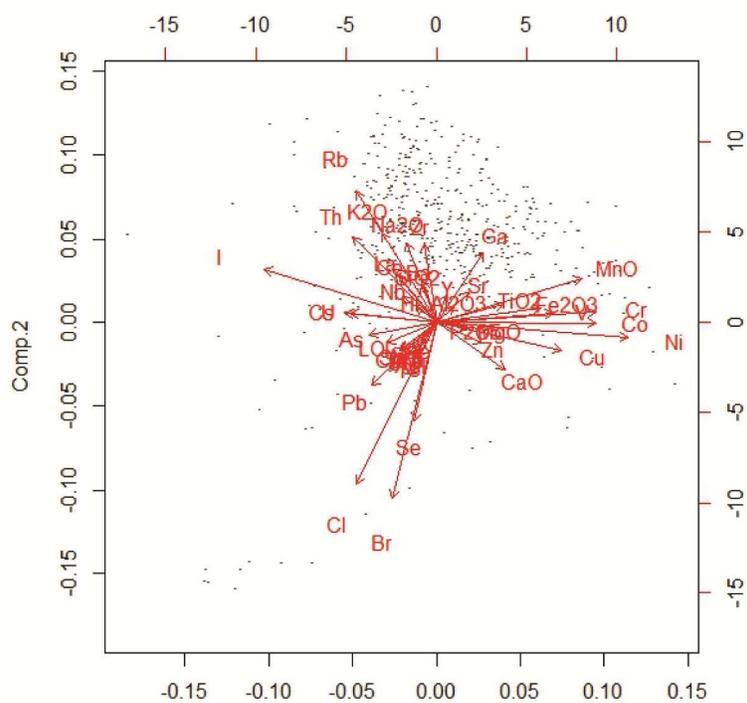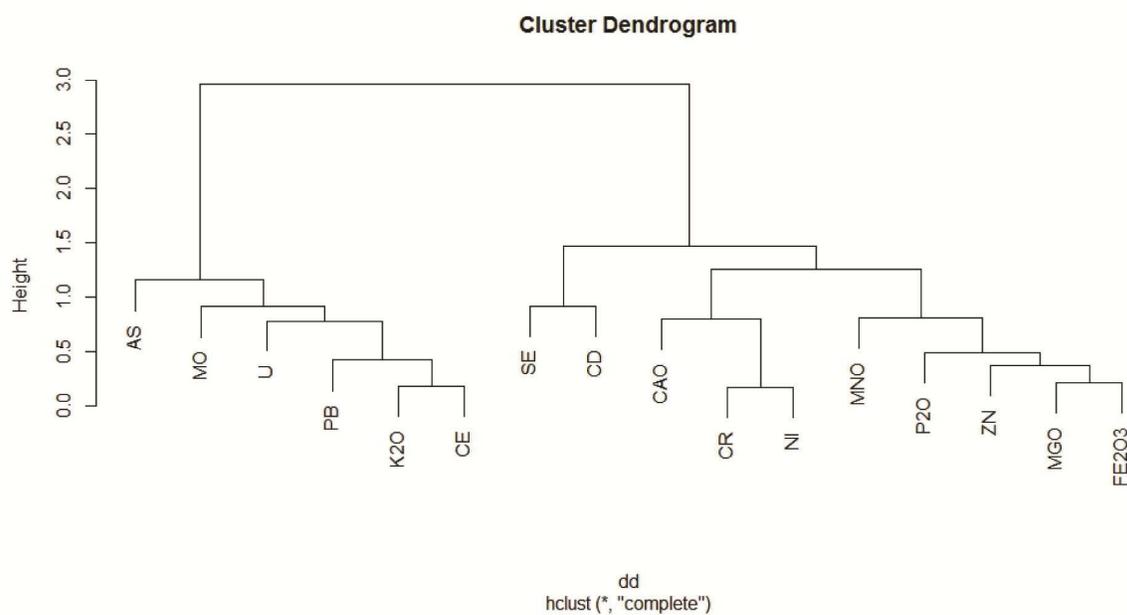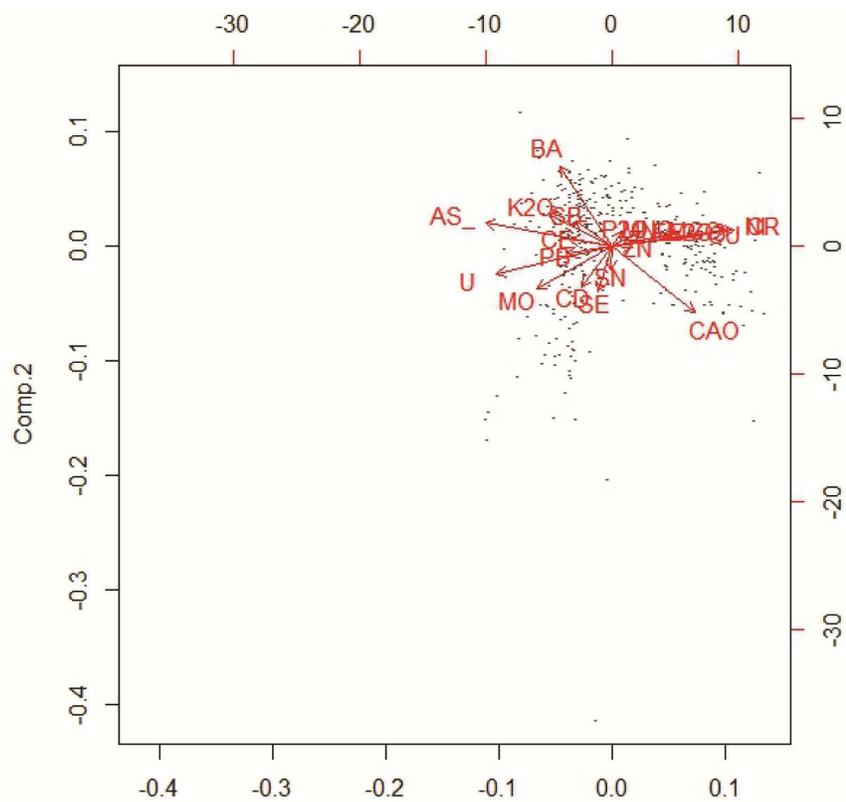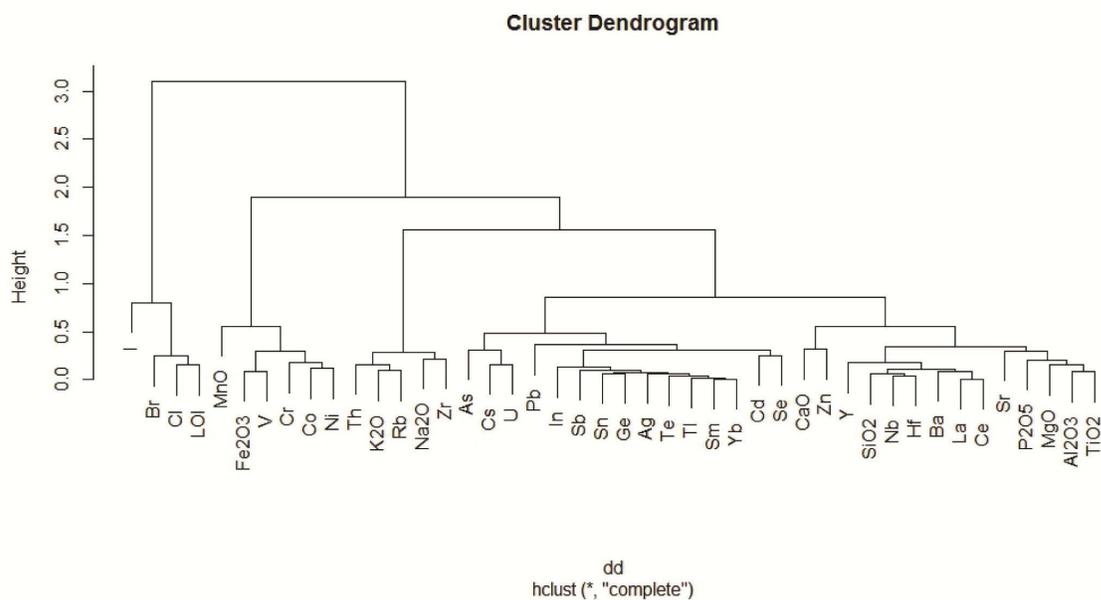
Highlights

- Incidences of Chronic Kidney Disease (CKD) are investigated for unknown aetiology.
- A compositional Poisson regression approach is demonstrated to investigate relations between CKD and environmental nephrotoxins including arsenic, cadmium, and lead.
- The validity of classical Poisson regression is discussed using compositional geochemistry covariates.
- Data driven and knowledge-driven log-ratio approaches are introduced.