Geospatial analysis of food composition data to estimate population micronutrient intake

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy
of the
University of London

July 2025

Lucia Segovia de la Revilla

PhD Student and Research Fellow in Nutrition Data

Department of Population Health

Faculty of Epidemiology & Population Health

London School of Hygiene & Tropical Medicine

With funding from Bill and Melinda Gates Foundation

Research group affiliations:

LSHTM Nutrition Group

Micronutrient Action Support Policy project

Abstract

Micronutrient deficiencies are widespread globally with highest prevalence in low- and middle-income countries. Spatial mapping of dietary micronutrient supplies and deficiency risks is important for identifying subnational variations that are often masked by national averages. This is particularly relevant for mineral micronutrients where geospatial variability of grain composition could influence the micronutrient intakes and status of the population. However, at present, food composition data are typically reported in national or regional food composition tables (FCTs), a structure that precludes the reporting and use of finer scaled, georeferenced food composition data. Hence, the aim of this thesis was to assess the implications of using georeferenced food composition data to estimate risk of inadequate intake in context with localised food systems.

First, a scoping review (Chapter 3) was conducted to assess the availability and geographic location of food composition data and its metadata for use in sub-Saharan Africa. In Chapter 4, a framework to compile reproducible, reusable, efficient and transparent food composition datasets, was developed. Using selenium in Malawi as case study, in Chapter 5, the different levels of spatial aggregation of high resolution (~250 m) maize selenium concentration data to estimate dietary selenium intakes was evaluated. A geospatial model was developed and applied comparing 10 levels of spatial aggregation to georeferenced (i.e. point data) plasma selenium concentration data among women (15-49 years old). Finally, in Chapter 6, the smallest level of spatial aggregation for maize selenium concentration (small area) and single maize values from national and regional food composition data were matched to apparent food intakes for dietary selenium intake estimations. The implications of using small area selenium content of maize and maize products vs single (national level) values from Malawi and from publicly available FCTs for estimating the risk of apparent inadequate selenium intakes was evaluated for Malawi.

This study's results highlighted the need for more geospatial information in food composition data, and the value of using high spatial resolution food composition data (including improved data systems), particularly when estimating sub-national apparent selenium intakes and inadequacy risks. The use of small area food composition data may increase the accuracy when identifying hotspots of micronutrient inadequacies which could help make informed decisions with regards to investments and resource allocation for nutrition surveillance, and the development and evaluation of interventions, particularly in contexts where dietary diversity is low and food systems are highly localised.

Supervisory team

Dr Edward Joy

Associate Professor in Nutrition and Sustainability

London School of Hygiene & Tropical Medicine

Dr Elaine L. Ferguson

Professor of Global Nutrition

London School of Hygiene & Tropical Medicine

Dr Claire Dooley

Lecturer in Spatial Data Science

Centre for Advanced Spatial Analysis

University College London

Declaration of own work

I, Lucia Segovia de la Revilla, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgement

Firstly, I would like to thank my supervisors for their guidance and support throughout this journey. Edward Joy, thank you for your patience, your useful comments and suggestions and, for helping me to see ‘the bigger picture’. Thank you, Elaine Ferguson, for your dedication and for sharing your infinite knowledge in the kindest possible way. And, thank you, Claire Dooley, for bringing joy to this PhD and for teaching me so much about a topic that has become so dear to me. Together, you made the best supervisory team a PhD student could have, not only because you are great experts in your field, but also because you are excellent human beings. I really hope you supervise many more students in the future!

I would also like to thank all my colleagues, particularly Rupinder Sahota, Fanny Sandalinas and Kevin Tang, with whom I shared more than an office. Thank you for the laughs, some rants, and for sharing so much knowledge and good advice. Thank you to the Nutrition Group for being always a source of expertise, and for operating as a supportive group. In particularly to Hilary Davies-Kershaw for her thoughtful feedback and overall kindness.

Thank you to those of you who were farther in space but, similarly close to my heart, specially to Louise Ander, for your moral support, compassion and for all the adventures this PhD had brought. Of course, I can’t forget, Tom Codd, Liberty Mlambo, Gareth Osman, and the extended MAPS ‘family’. Thank you for your technical advice and insightful discussions, for putting context to the data and, in general, for all the support and fun you brought to this project. You were part of this journey, and I am very grateful for your help and companionship.

Thank you to Fernanda Grande, Doris Rittenschober, Ana Moltedo, Bridget Holmes and the team at FAO for sharing your in-depth knowledge and ideas, your passion for food composition and for thoughtfully contributing to this thesis. It has been a pleasure to work with you.

Thank you also to all my friends, particularly all the strong women who have share all or part of the journey with me. You are an inspiration and a source of motivation to keep moving forward. A special, thanks to Janet Molina Maturano for all the encouragement and for holding my hand in this journey. And, to Filipa Soares for always being comforting and reassuring, and for inspiring me to achieve great things. Gracias sisters!

Now it is time to thank my family. Gracias a mi familia por creer en mí y por animarme siempre. En especial, quiero dar las gracias a mis padres por ser únicos, geniales y por darme todo sin pedir nada a cambio. A mi madre por su amor incondicional, por estar siempre ahí y quererme siempre más. A mi padre por apoyarme y ayudarme siempre. A mi hermana, porque tu sí que eres la mejor hermana que se puede tener. Gracias Carmen por escuchar mis quejas y siempre, siempre hacerme reír. Y a Javi, mi compañero, who can care less about this PhD, pero que lo ha sufrido igual.

Esta tesis se la dedico a mis abuelos, a mi abuela Carmelin, a mi abuelo Antonio, a Queta y al Jefe. Os quiero.

Table of Contents

Abbreviations

ABBREVIATION	DEFINITION
AFE	Adult Female Equivalent
AGP	α-1 acid glycoprotein
Ca	Calcium
COVID-19	Coronavirus Disease
CRP	C-Reactive Protein
DHS	Demographic and Health Survey
DIC	Deviance Information Criteria
DIO	Iodothyrosnine deiodinases
EA	Enumeration Area
ESN	FAO Food and Nutrition Division
FAO	Food and Agriculture Organization of the United Nations
FBS	Food Balance Sheets
FCT	Food Composition Table and Databases
Fe	Iron
GPS	Global Positioning System
GPX	Glutathione Peroxidase
HCES	Household Consumption and Expenditure Survey
I	Iodine
ICP-MS	Inductively Coupled Plasma-Mass Spectrometry
IHS4	Fourth Integrated Household Survey
IHS5	Malawi Fifth Integrated Household Survey
IML	International Food Composition Tables
INFOODS	International Network of Food Data Systems
INLA	Integrated Nested Laplace Approximation
LSHTM	London School of Hygiene and Tropical Medicine
LSMS	Living Standards and Measurement Study
MAPS	Micronutrient Action Policy Support

USDA United States Department of Agriculture

WFDA WorldFood Dietary Assessment System

Zn Zinc

1 Chapter 1: Background

1.1 Mineral micronutrient status

Mineral micronutrient deficiencies are affecting more than two billion people worldwide, with children and women at particularly high risk (Black et al., 2008; Stevens et al., 2022). For example, the prevalence of iron (Fe), selenium (Se) and zinc (Zn) deficiencies among woman (15-49 years old) in the 2015-2016 Malawi micronutrient survey (2015-16 MDHS-MNS) were 15%, 63% and 63% respectively, based on biomarkers of status (National Statistical Office (NSO) [Malawi] & ICF, 2017; Phiri et al., 2019).

These deficiencies have far-reaching consequences, extending beyond individual health to impact entire societies. Fe deficiency is one of the main causes of anaemia and is linked with disruptions in the endocrine and immune systems. It is also key during pregnancy, when women are at the highest risk, leading to poor birth outcomes (e.g. low birth weight) and other perinatal complications. Zn deficiency is also associated with immune system impairment, poor growth and increased mortality risk, while Se deficiency is related to cardiovascular and thyroid disorders. In summary, maternal and child micronutrient deficiencies increase the risk of impaired growth and cognitive development, perinatal complications and increase the risk of morbidity and mortality. This could have long lasting effects impacting health, wellbeing and productivity in later life, contributing to a vicious cycle of poverty and malnutrition (Bailey et al., 2015; Black et al., 2013; Brown et al., 2021).

Although widespread globally, the highest prevalence of mineral micronutrient deficiencies is found in low- and low-middle-income countries. In these settings, multiple micronutrient deficiencies often occur simultaneously and appear to be driven by a common cause: inadequate dietary micronutrient intake. This inadequate dietary intake is common in sub-Saharan Africa where dietary diversity is typically low and, in some countries, reliance on localised food systems is high (Beal et al., 2017). Therefore, dietary nutrition data is needed to evaluate population micronutrient intakes, to identify sub-populations at risk of inadequate intakes and their main food sources of nutrients. Such data make an important contribution to the design, monitoring and evaluation of cost-efficient programmes that target one of the root causes of the problem (Beal et al., 2017; Ingenbleek et al., 2017; Joy et al., 2014; Neufeld et al., 2017).

1.2 Measuring mineral micronutrient intakes

The estimation of population micronutrient intakes relies on the availability of relevant food consumption and food composition data. These two datasets are matched, and nutrient intakes are calculated as the sum of products between quantity of the food consumed by the population and the nutrient composition of the foods (Eq.1). These datasets are required for estimating the percentage of population at risk of inadequate nutrient intakes, designing food-based guidelines or recommendations, planning food fortification interventions or evaluating whether local food systems will cover a population's nutrient requirements when intakes are reliant on it. They are also underlying data sources for numerous indexes (e.g. food security, undernutrition, cost of the diets, etc.) which are used for monitoring and evaluation of the nutrition targets and Sustainable Development Goals (SDGs). Hence, the veracity of the results are directly related to the accuracy and reliability of the consumption and nutrient composition data used (Beal et al., 2024; Coates et al., 2012; Micha et al., 2018).

$$Nutrient\ (K)\ intake_{d,i} = \sum_{f=1} \left[Nutrient\ (k)\ content_{100g}^{(f)} x \frac{Consumed\ food\ amount^{(f),d,i}}{100} \right] \qquad \text{(Eq.1)}$$

Where K is the nutrient intake by the individual (i) per day (d). And, k is the nutrient content in 100g of fresh weight edible portion in each food (f) consumed (in g) by the individual (i) per day (d).

1.2.1 Food consumption data

The best method to determine food consumption are individual-level dietary data (e.g. weighed or estimated food diaries, 24h – recalls) however, these methods are extremely resource intensive. Hence, they are often collected in surveys with small population sample sizes (Ferguson et al., 1989; Gibson & Ferguson, 2008; Leclercq et al., 2019). Thus, when resources are limited, studies evaluating micronutrient intakes at large scales (e.g. national) have used proxy dietary data (e.g. Food Balance Sheets (FBS)) matched with food composition data (Arsenault et al., 2020; Ferguson et al., 1989; Gibson and Ferguson, 2008; Joy et al., 2014). Recently, household survey data (e.g. Household Consumption and Expenditure Survey (HCES) data) has increasingly been used by the nutrition community as a source of dietary data to evaluate micronutrient apparent intakes of populations (Tang et al., 2022; World Health Organization et al., 2021).

The World Bank Living Standards and Measurement Study (LSMS) supports a family of HCES with a broadly consistent scope and design. In these HCES, a food consumption module is included in which data are collected on food consumed by the whole household, over a period of seven to 14 days prior to the questionnaire administration. Typically, HCES surveys run on a five-yearly cycle and aim to provide national coverage. Besides the food consumption module, HCES also collects a wide range of other information, including household socioeconomic characteristics and displaced geographic location of the household allowing for spatial analysis. Previous studies have used the Malawi Third and Fourth Integrated Household Survey (Malawi 2010-11 IHS3 and Malawi 2016-17 IHS4, respectively) to evaluate the micronutrient apparent intakes of the Malawian population (Joy et al., 2015; Tang et al., 2021). In this study, the Malawi 2016-17 IHS4 was used despite the availability of a more recent wave: The Malawi 2019-20 IHS5. The IHS5 was collected during COVID-19 which may have introduced some bias, including the impact of the lockdown on the data collection (e.g. some households were not surveyed). Furthermore, the latest 2015-16 Malawi Micronutrient Survey (2015-16 MDHS-MNS) carried as part of the Demographic and Health Survey (DHS), used in this study, was closest in time to the Malawi 2016-17 IHS4.

1.3 Mineral micronutrient content in foods

1.3.1 Food Composition Tables and Databases

Food composition tables and database (FCTs), in addition to other information, provide data on the energy, water and nutrient content of foods (expressed per 100 grams edible portion). Beside the commonly known uses outlined in previous sections, particularly when combined with food consumption or acquisition data to estimate the nutrient intakes of population, and for informing policy decisions on public health nutrition, they are also used in food labelling and product reformulation to meet government dietary recommendations (Traka et al., 2020). Nutrient content of foods can be used to guide agricultural policies aiming at providing more nutritious foods, by identifying crops varieties with higher micronutrient content, and for assessing the efficacy of biofortification practices (i.e. applying micronutrient-rich fertilisers to crops) (Charrondière et al., 2013). They could be used to monitor changes in nutrient supplies over time, for instance, changes due to regulatory practices (e.g. reduction in sugar content) or food trade agreements (Broadley et al., 2006; Hutchinson et al., 2021). Despite the importance of these datasets, the attention in their collection, collation, standardisation and overall quality is often insufficient.

The micronutrient content variability in foods is high, particularly for mineral micronutrients for which their concentration can vary not only due to variety, maturity and processing but also due to environmental factors, such as soil pH (Charrondière et al., 2013; Dunlop et al., 2025; Greenfield & Southgate, 2003). For instance, the Ca concentration in teff in Ethiopia ranged from 46.8-7,935 mg kg^{-1} while the Se concentration in maize in Malawi ranged from 1-1,788 µg kg^{-1} (Gashu et al., 2021). To capture this variability in composition (and therefore uncertainty in resultant estimates of dietary nutrient intakes), ideally multiple georeferenced samples of food would be collected from different locations and analysed to support the development of subnational and/or small area, preferably georeferenced, FCTs. However, the collection, preparation and laboratory analyses of food samples for compositional analysis is costly, e.g. >USD3,000 per food item for complete analysis (Dunlop et al., 2025; Vincent et al., 2020). So, rather than analysing multiple individual samples, analysis of composite food samples is common, which aims to represent the 'average' nutrient content of that food in the specific FCT context. For example, by collecting multiple samples of *injera* in Ethiopia and blending them together, the data obtained after analysis could be seen as the average of the nutrient content of *injera* in Ethiopia. Although this method is cost-effective, it fails to account for the variation in nutrient content in foods. Alternatively, nutrient values in FCTs are sourced and imputed from published sources (e.g. peer-reviewed manuscripts, other FCTs, etc.) and sometimes the only available data are from other geographies. For example, 38% of the datapoints in the recently published Malawi FCT (2019) were borrowed from international FCTs and there are a high proportion of missing values for Se and iodine (I) (van Graan, et al., 2019). All of these techniques are often combined when producing FCTs which difficult the study of the nutrient variability and its potential implications for human intakes.

Currently, the level of reporting and metadata of FCTs varies widely, from very informative, machine-readable metadata found in the USDA FCT (2019) to very limited information in the Malawi FCT (2019). In the Malawi FCT, metadata (i.e. information about the nutrient data points in the FCTs) is not available and the references are reported at the food item level, rather than food item-nutrient level. Practitioners using this FCT would be unable to identify key information such as, sample location, analytical methods, and age of the data for each nutrient value. This information (metadata) is essential for understanding and/or quantifying the extent and impact of local food samples in FCTs, particularly when developed to assess and inform populations that are heavily reliant on localised food systems.

Part of the problem is that guidelines and principles of FCT compilation were developed more than one decade ago (FAO/INFOODS, 2012b, 2012a; Greenfield & Southgate, 2003). Despite the usefulness of the information provided, the current progress in data management and architecture principles as well as computing power make these guidelines largely obsolete. For instance, following advances in other fields, modelling of nutrient concentration in some foods and/or food groups have been attempted. Hicks et al. (2019) estimated the nutrient supplies of marine fish using a modelling approach and published analytical values. Similarly, Gashu et al. (2019) surveyed crop samples across Ethiopia and Malawi producing estimates of mineral concentration (and their uncertainty) for every square grid in Malawi and Ethiopia (except in masked areas). However, these advanced techniques have not yet been integrated in the collation of FCTs and/or used to support subnational estimation of mineral food composition for human nutrition.

The lack of good quality primary food composition data and inadequate reporting of metadata (including (GPS) location of the samples) is problematic, and may be the reason for the low uptake, when considering the use of modelling approaches to enhance food composition data. Hence, new data structures and standards should be put forward which would allow the collation of georeferenced, metadata-rich data points for use in nutrition. These metadata-rich FCTs, which would better reflect the underlying sources of nutrient variability, would help to accurately quantify and understand the prevalence and aetiology of micronutrient deficiencies (Gashu et al., 2021; Joy et al., 2015; Ligowe et al., 2020).

1.3.2 Compiling Nutrient Conversion Tables for dietary assessment

The nutrient information provided in the FCTs is compiled into Nutrient Conversion Tables when used for assessing population nutrient intakes. The NCTs are survey-specific datasets that contain the energy and nutrient content of the food reported as consumed in the specific survey (i.e. the Se content of the food listed as consumed in the Malawi 2016-2017 IHS4). They are often compiled from more than one FCT, and it could include nutrient values from other sources (e.g. published reports, analytical data, etc.). When the nutrient data sources used to compile these NCTs, primary publicly available FCTs, are of insufficient quality and/or lack adequate metadata, as shown in previous sections, these uncertainties will be propagated to the NCTs, and subsequent dietary intake estimates.

The accuracy of the food intake has been the focus of the nutrition agenda in the past decade, with many initiatives working to increase the availability, accuracy, representativeness and granularity of the dietary data (Coates et al., 2012, 2017; Gibson & Ferguson, 2008; Leclercq et al., 2019). For example, the World Health Organization (WHO)/ Food and Agriculture Organization (FAO) of the United Nations have recently published multiple individual-level dietary data in their GIFT platform. In addition, the FAO together with Intake, an organisation that aims to increase the availability, quality, comparability, and use of reliable dietary data and metrics, have been partnering with government in diverse countries to increase the frequency and scope of their national dietary data collection (FAO & Intake, 2022). Unfortunately, the other side of the equation (Eq.1), the nutrient content in foods has received less attention, particularly in low- and middle-income countries, where only a handful of new FCTs have been developed, and even less primary data have been collected (Bruyn et al., 2016).

The lack of local mineral composition data would affect the reliability and representativeness of the mineral intake estimates produced using ‘borrowed’ values from other countries. This is particularly problematic in contexts where highly localised food systems and monotonous diets are the main source of most of the mineral micronutrients.

1.3.2.1 Localised food systems

The term ‘local food system’ refers to food systems where foods are mainly sourced from local farmers/producers. However, due to differences in farming/production systems, the scale at which these local farmers/producers operate (e.g. national or village level) can differ depending on the context and, thus, have different implications when estimating micronutrient intakes (FAO & WFP, 2018; Nguyen & Qaim, 2025).

In this thesis, our analysis focuses on Malawi, where smallholder farmers cultivate 70% of arable land and are responsible of 75% of the crop production. Most of these farmers grow food crops, with maize being the main crop and staple food in the country, which incentivised by agricultural policies, have contributed to the low dietary diversity of the population, as highlighted by the National Agriculture Policy (Government of Malawi, 2024).

1.3.2.2 Monotonous diets

Low dietary diversity is often associated with higher risk of insufficient (mineral) micronutrient intakes, as populations are highly reliant on a few crops for a large proportion of their mineral micronutrient intakes. For example, in Malawi, maize is consumed by >90% of households, and provides >50% of caloric intake, and more than 30%, 40%, and 20% of Fe, Zn and Se dietary supplies at national level (Joy, Kumssa, et al., 2015). Similarly, maize is highly consumed in most of the countries in Southern and Eastern Africa such as Lesotho, Zambia, Zimbabwe and South Africa (Galani et al., 2020). Therefore, variation in the mineral concentration in maize would have an impact on the intakes of these populations.

The combination of localised food systems and monotonous diets are frequently seen in context where high proportion of the population is engaged in agriculture and where agricultural policies have traditionally focused on one (or very few) staple crop. As it is the case of Malawi, as well as other countries in the region (Zimbabwe, Uganda, Tanzania), where the main crop is maize, and most of the crop production comes from smallholder farmers, contributing to the low dietary diversity in these countries (Galani et al., 2020; Nguyen & Qaim, 2025).

This, in turn, could be aggravated by the strong spatial variability of the mineral micronutrient content in main staple crops in sub-Saharan Africa. This spatial variability can affect the nutrient intake, as the mineral concentration in maize would differ by its growing location, with values varying in up to 2 orders of magnitude (Gashu et al., 2021).

1.4 Selenium as a case study

Selenium is an essential mineral micronutrient and is a constituent of multiple selenoproteins with diverse functions in the human body, such as glutathione peroxidase (GPX), selenoproteins P (SEPP), iodothyrosnine deiodinases (DIOs), etc. Selenocysteine (SeCys) is the functional form of the Se in these selenoproteins. The main functions of these selenoproteins are related to the inflammatory response, thyroid regulation, and fertility (Naderi et al., 2021; Rayman, 2020). Due to the role of Se in reducing oxidative stress and inflammation response, a number of synthetic Se compound have been developed and tested such as Ebselen or Diphenyl diselenide to prevent and treat various diseases (e.g. atherosclerosis, strokes or cancer) (Naderi et al., 2021).

Two endemic diseases have been described in relation to severe Se deficiency: Keshan disease and Kashin-Beck disease. The former is a myocarditis which affects young children (2-10 years old) while the latter is an osteoarthropathy. Both were first described in certain areas of China and Russia with extremely low Se levels in foods available in their food systems. Se deficiency also has been associated with thyroid autoimmune diseases and other conditions (Rayman, 2020). More recently, dietary Se intakes have been associated with the composition of the gut microbiome (Bhatnagar et al., 2024; Rayman, 2020). However, excessive intakes of Se can lead to toxicity (acute or chronic selenosis), and some studies suggest it may increase the risk of type 2 diabetes although the evidence remains inconclusive (Ma et al., 2024; Naderi et al., 2021). Hence, due to its ‘U shape’ relationship between the dose and the health benefit, it is necessary that measurement of dietary intakes are as accurate as possible.

1.4.1 Assessment of the selenium status in populations

The Se status is referred to as the Se available in the human body for metabolic functions and it is a function of the Se intakes, body pools of functional Se, and Se excretion. Selenium is found in a wide range of selenoproteins which all have in common a selenocysteine amino acid. The Selenoprotein P (SelP) is the most common selenoprotein in plasma and is key in the transportation, storage and metabolism of Se in the human body.

A number of tissues have been used to quantify Se status in humans: hair, toenails, whole blood, blood plasma and urine. They are all linked to dietary Se intakes however they may reflect different aspect of status. For example, toenails and hair represent long-term exposure to Se, and also are more prone to contamination from non-dietary sources, such as using Se containing shampoos. Urine is a good biomarker for excretion, and it has been shown to correlate well with plasma and it can show spatial patterns in contexts with localised food systems, hence showing some potential for population-level monitoring (Noisel et al., 2014; Phiri et al., 2020). Furthermore, Se status can be measured through the selenoprotein activity, such as GPX3 in plasma or GPX1 in whole blood/erythrocytes. However, when Se concentrations are sufficient, their activity would plateau due to saturation and hence it is not recommended in contexts were low and high Se concentrations are likely to be found in the population, such as in Malawi (Combs, 2015; Stefanowicz et al., 2013).

In this study, plasma Se concentration was used as it is the recommended biomarker for assessing population status due to its responsiveness to Se intakes, i.e. plasma Se is primarily determined by Se intakes. It can be effectively measured using Inductively Coupled Plasma Mass Spectrometry (ICP-MS). Beside intakes, there are other factors that can influence plasma Se concentration at the individual level, such as gender, age, pregnancy and lactation, smoking status, some diseases and inflammation. In our analysis, only age and inflammation (C-Reactive Protein (CRP) and α-1 acid glycoprotein (AGP) concentrations as continuous variables) were considered because, only non-pregnant women (15-49 years old) were included in our analysis. Lactation information was not available and smoking prevalence was <1% for women (15-49 years old) in Malawi (National Statistical Office (NSO) [Malawi] & ICF, 2017).

Several cut-off values are used to identify Se deficiency, depending on the clinical end-point and different functions in the body (Figure 3). Although the plasma Se concentration that prevents the development of Keshan disease is very low, the threshold that it is commonly used to define a healthy individual is <84.9 ng mL^{-1} which is the concentration of Se in plasma at which GPX3 plateaus. Hence, that is the cut-off used in this study to assess risk of deficiency in women (15-49 years old) in Malawi (Phiri et al., 2019; Thomson, 2004).

1.4.1.1 Dietary intakes

The dietary Se intake is the main driver of Se status, and it can be estimated by measuring food intakes and the Se concentration in foods. The linkages between the bioavailable Se in soil and its concentration in the food systems have been described previously, meaning that populations that live and obtain their foods from low plant-available Se areas would likely have low dietary Se intake and, in turn, low Se status and vice versa (Hailu et al., 2023; Hurst et al., 2013). In the Eastern and Southern regions of Africa, most of the soils are considered low in (plant-available) Se with small pockets of high and very high Se (Gashu et al., 2021; Ligowe et al., 2020; Ngigi et al., 2020). Furthermore, populations in these areas are heavily dependent on cereals and other staple crops for their Se supplies paired with low consumption of animal-source foods. Assessment of dietary Se intakes and sources is important in these contexts to identify populations that could be at risk of inadequate Se intakes, and to plan, monitor and evaluate interventions to alleviate their potential Se deficiency.

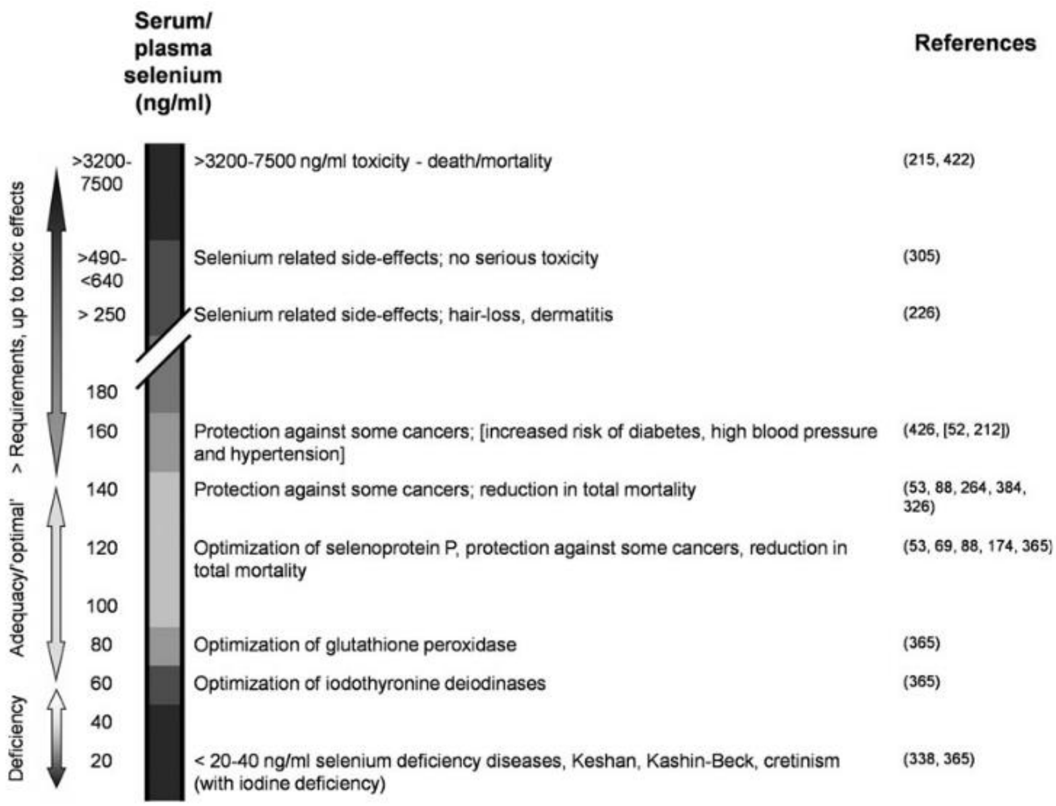


Figure2. Relationship between plasma selenium concentration and health-related effect (Source: Fairweather-Tait et al., 2011).

2 Chapter 2: Thesis overview

2.1 Thesis Road Map

This thesis is organised in seven chapters: **Chapter one** provides the background information of this thesis. Here, in **Chapter two**, the thesis overview is presented, including, this road map, evidence gaps and rationale. It also outlines the aim and objectives, overall methods and timelines and funding. **Chapter three** is a scoping review that was conducted to identify published food composition datasets (FCTs/NCTs) used for dietary assessment of the mineral micronutrients calcium (Ca), Fe, I, Se and Zn in sub-Saharan Africa for frequently consumed food items. **Chapter four** presents a study where an open science framework and tools to compile reproducible food composition data for use in nutrition was developed. **Chapter five** is a spatial modelling study that evaluated different level of aggregation of high-resolution maize grain Se concentration data in Malawi. It was assessed by modelling the spatial relation of plasma Se concentration in non-pregnant women (15-49 years old) and maize grain Se concentration in Malawi at different aggregation levels. **Chapter six** calculate and compare the dietary Se estimates and percentage of the population at risk of inadequate Se intakes in Malawi under three scenarios. The three scenarios were: a) baseline scenario used current practices in dietary assessment of using publicly available FCTs for collating Se concentration in foods, b) national-single maize scenario replaced maize and maize flour Se content by one single national average value, and c) small area maize scenario used the smallest aggregation, based on household Enumeration Areas (EAs), identified in chapter five to calculate the Se concentration in maize and maize flours. **Chapter seven** discuss the thesis' findings, strengths and limitations and the ways forward and closes the thesis with the overall conclusion.

2.2 The rationale

2.2.1 Spatial variation in mineral micronutrient composition

Several studies have reported spatial patterns in mineral micronutrients in staple crops which means that the mineral content in some foods may vary across different locations in sub-Saharan Africa (Belay et al., 2020; Gashu et al., 2021; Kihara et al., 2020; Phiri et al., 2019). The mineral content of these crops are greatly influenced by agro-environmental factors such as soil pH, organic matter, inputs and fertilisers, crop type, etc (Botoman et al., 2022; Chilimba et al., 2011; Greenfield & Southgate, 2003; Hailu et al., 2015). These factors lead to nutritionally relevant spatial variation in the mineral micronutrient content in crops, as reported for the composition of Ca, Fe, Se, Zn amongst other micronutrients in maize grain in Malawi, the Kilimanjaro District in Tanzania and in several districts in Western Kenya and for green leafy vegetables for calcareous and non-calcareous soils in Malawi (Joy, et al., 2015; Watts et al., 2019). More recently, Gashu and colleagues (2021) also reported nutritionally important spatial variation in the composition of Ca, Fe, Se and Zn in teff and wheat in Ethiopia and in maize grain in Malawi. The spatial pattern in the variation of maize grain Ca, Fe and Se concentration occurs over distances of up to 50–80 km in Malawi (Gashu et al., 2021). These patterns are illustrated in Figure 1a, whereby grain concentrations from sites close together are typically more similar than those far apart, showing a clustered distribution whereas Figure 1b illustrates a situation with uniform pattern. These studies illustrate the importance of studying spatial patterns in food composition data, including for Se which show strong geographic variation. However, to date, there is no georeferenced food composition database (i.e. a database with data on food samples, their food components and the sampling location (e.g. coordinates)), partly due to the cost of collecting and analysing food samples, and for Se, the analytical challenges of measuring their very low concentrations, particularly in maize grown in areas where plant-available Se is deficient (Kumssa et al., 2022; Muleya et al., 2021).

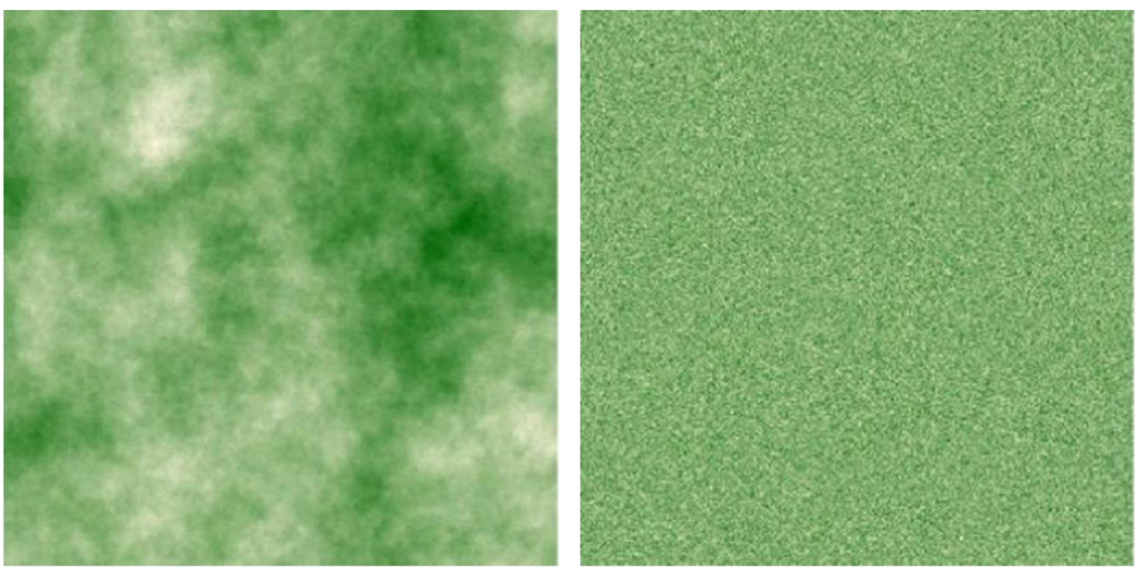


Figure 1a Figure 1b

Figure 1: Illustrative map of grain micronutrient concentration (low=white, high=dark green) with clustered (1a) and uniform (1b) distribution. The two maps have the same mean and standard deviation concentrations. Source: R. Murray Lark, personal communication.

Accounting for geospatial variation in the mineral content of food is particularly important in countries or regions in sub-Saharan Africa, where dietary diversity is typically low, diets are dominated by predominantly locally produced cereals or other staple crops (e.g. maize, rice, *teff,* or cassava) (Bergau et al., 2022; Galani et al., 2020; Nguyen & Qaim, 2025). For instance, maize, which is often home-produced or sourced locally, contributes around 50% of total energy in Malawian diets and is the main dietary source of Se (Government of Malawi, 2024; Joy, Kumssa, et al., 2015). Thus, the variation in the Se content of maize, which can be up to 10-fold, would substantially affect Se intakes, depending on the area of residence (Chilimba et al., 2011; Gashu et al., 2021; Joy et al., 2015).

In summary, Se concentration in maize in Malawi was chosen as a case study because: 1) maize is widely consumed in Malawi by >90% of the population, 2) the Se spatial variability in soil, crops and dietary intake has been extensively documented (Chilimba et al., 2011; Hurst et al., 2013; Joy, Kumssa, et al., 2015; Ligowe et al., 2020), 3) the spatial variability in crop Se concentrations are nutritionally-relevant and occur at distances that will likely influence dietary Se intake and the dietary Se status of populations at subnational levels (Gashu et al., 2021; Joy, et al., 2015) (Figure 1a), and 4) plasma Se is responsive to Se intake (Combs, 2015; Fairweather-Tait et al., 2011; Hurst et al., 2013; Ligowe et al., 2020).

2.3 The evidence gap

Existing evidence shows that the mineral content of foods is geospatially variable depending on local environmental conditions. This variation has the potential to influence dietary mineral intakes for population that are reliant on locally grown staple crops as described in Chapter 1 (Abdu et al., 2022; Hailu et al., 2015; Joy, Kumssa, et al., 2015; Ngigi et al., 2020; Watts et al., 2019). Hence, mineral composition data and metadata in FCTs, including the geographic origin of the data reported, for assessing dietary mineral intakes in sub-Saharan Africa is crucial. This information is needed to critically understand the variation in mineral inadequacy risks, particularly when there is spatial clustering, and the bias it might be introduced when local mineral composition is not available. However, FCTs generally do not provide geographically specific mineral composition data for foods, and the extent to which the data and metadata are available, on the mineral content of crops grown in sub-Saharan Africa, particularly for Se, is not known.

Recently, georeferenced crop samples have been collected to provide representativeness of the mineral composition in crops grown in Malawi (and Ethiopia) (Gashu et al., 2021; Kumssa et al., 2022), however, there is no guidance on how to use these georeferenced point datasets for dietary assessment. For instance, the level of aggregation of maize Se concentration (i.e. at cluster level or at district level), which would represent the 'average' Se content in maize consumed by population in Malawi is not known. To date, only a few studies have collected georeferenced samples and analysed their mineral composition for assessing dietary mineral intakes (Joy, Kumssa, et al., 2015; Watts et al., 2019). These studies collected small number of georeferenced food samples (e.g. using convenience sampling) which were not geographically representative, hence the mineral composition data used could only be aggregated into two levels based on soil type (calcareous and non-calcareous). Hence, when estimating dietary Se intakes in Malawi, the level of aggregation of maize Se concentration that would be more representative of the maize consumed in the country is yet to be known.

Finally, despite the well-known geographic variation in Se concentration in foods, the extent to which using e.g. a national average single value of maize composition for the whole country could under- or over-estimate the dietary Se intake and mask important subnational differences is unknown.

2.4 Aim and objectives

The hypothesis was that the use of small area (i.e. smaller than country regions) food composition data derived from georeferenced sampled data for dietary assessment will increase the accuracy of estimating the percentage of the population at risk of inadequate intakes particularly at subnational levels. This is because, small area estimates will reveal subnational differences in mineral composition that would be otherwise masked by national (single point) composition data. The **aim** of this thesis was to assess the implications of using small area food composition data to estimate risk of inadequate intake in context with localised food systems and the main question of this thesis was '*What are the implications of using small area food composition data to estimate risk of inadequate intake in sub-Saharan Africa?*' To answer our main question, a series of questions related to the objectives of our study are proposed.

2.4.1 Research questions

1. What is the data availability and the geographic scope of existing food composition data (including FCTs & NCTs) available to estimate dietary micronutrient (Ca, Fe, I, Zn, and Se) intakes in sub-Saharan Africa?
   1.1. For selected food items, what is the availability of mineral micronutrient composition data in sub-Saharan Africa FCTs/NCTs (i.e. available vs missing for each food item and mineral combination)?
   1.2. For available data in the sub-Saharan Africa FCTs/NCTs, what is the information on geographic location and the level of reporting (i.e. at food item level, or at food item and mineral combination level)?

2. What is the suitable level of spatial aggregation of georeferenced crop composition data to estimate the percentage of populations at risk of inadequate intakes of Se in Malawi?

2.1. Provided that maize is the main staple in Malawi, what is the level of spatial aggregation that provides the highest association between maize grain Se concentration and plasma Se concentration in women (15-49 years old) in Malawi?
2.2. What socio-demographic parameters (i.e. wealth quintile, household location (rural vs urban)) affect/influence the association between maize Se concentration and plasma Se concentration in women (15-49 years old) in Malawi?

3. What are the implications of using small area (i.e. based on household EAs) crop composition data to estimate the percentage of populations at risk of inadequate intakes of Se in Malawi?
   3.1. What are the differences in the estimates of dietary Se intakes of the population in Malawi by residency and at subnational level when using small area crop composition estimates compared with national or regional food composition data?
   3.2. What are the differences in the estimates of the percentage of populations at risk of inadequate Se intakes when using small area crop composition estimates compared with national or regional food composition data?
   3.3. What recommendations can be made about the use of small area crop composition estimates and its integration with dietary datasets in other settings?

Following the findings and reporting of the scoping review (Chapter 3), an opportunity arose to partner with the FAO to develop a framework (Chapter 4) aimed at closing data and metadata gaps identified in the scoping review, particularly the lack of transparency of food composition data as used for dietary assessment. Hence, the objectives of the thesis were expanded to include the development and application of this framework.

2.4.2 Objectives

The objectives of this study were:

i) to assess the current availability of mineral food composition data and metadata (including geographic location) for estimating dietary micronutrient supplies in sub-Saharan Africa.

ii) to develop a framework for generating and reporting transparent and reproducible FCTs/NCTs for use in nutrition research.

iii) to identify the suitable level of aggregation of mineral composition data derived from georeferenced crop samples for estimating dietary nutrient intakes of populations when high resolution food composition data is available.

iv) to evaluate the potential implications for public health policy decision-making of using small area maize Se composition data when quantifying dietary Se intakes, against the current practice of using national level (i.e. one single value of maize Se concentration) food composition data.

2.5 Methods

The methods used in this thesis are described in each corresponding chapter (Chapters 3-6). In addition, a repository for each chapter methods and analysis are available on my [GitHub](https://github.com/LuciaSegovia).

2.5.1 Data sets and data management

In line with good data management practices a data management plan was developed and maintained, and it is attached (Annex 1). A summary of the datasets used in this study is provided in Table 1.

Table 1. Summary of the datasets used in this thesis, and information on the spatial scope and data provenance.

Datasets	Spatial information	Data source	Citation
Georeferenced maize selenium data[1]	Point location (GPS coordinate)	GeoNutrition project	Kumssa et al. (2021)
Georeferenced maize selenium data	Point location (GPS coordinate)	Chilimba et al., (2011)	Chilimba et al. (2011)
Georeferenced socio-economic information of women (15-49 years old) in Malawi.	Cluster point location (displaced GPS coordinates)	2015-16 Malawi Demographic and Health Survey (2015-16 MDHS)	National Statistical Office and ICF. (2016a)
Georeferenced plasma Se concentration and other individual-level information on women (15-49 years old) in Malawi.	Cluster point location (displaced GPS coordinates)	2015-16 MDHS Micronutrient Survey (2015-16 MDHS-MNS)	National Statistical Office and ICF. (2016b)
Georeferenced household food consumption and other household information in Malawi	Cluster point location (displaced GPS coordinates)	Malawi Fourth Integrated Household Survey, 2016-2017(Malawi 2016-17 IHS4)	National Statistical Office. (2017)
Se concentration in foods	National (one single point)	Malawian FCT (2019)	van Graan, et al. (2019)
Se concentration in foods	National (one single point)	Kenya FCT (2018)	FAO & Government of Kenya. (2018).
Se concentration in foods	National (one single point)	Lesotho FCT (2006)	Lephole, et al. (2006)
Se concentration in foods	National (one single point)	UK FCT (2021)	Public Health England. (2021)
Se concentration in foods	National (one single point)	USDA FCT (2019)	United States Department of Agriculture (USDA). (2019)

[1]This data was also published in Gashu et al (2019). It has national coverage. Global Positioning System (GPS)

2.6 Ethics

The LSHTM ethical approval (Ref. 26546) was granted on 9th May 2022, and it is attached (Annex 2). Furthermore, all the data used in this thesis is secondary data which have been anonymised and de-identified by the data owners prior access. Research ethics training was completed in the 5th July 2021 as part of the doctoral training plan (Annex 3).

2.7 PhD timeline and Funding

This PhD was conducted part-time as staff member of the LSHTM from February 2021 to December 2024. I successfully upgraded in February 2022. I developed the open science framework in 2022, performed the spatial analysis in 2023, and studied the implications of using georeferenced maize Se composition for dietary assessment in 2024. I finalised the thesis writing in early 2025. As part of this thesis, I have published Chapter 3 and 4 in peer-reviewed journals and presented my work in numerous conferences (Annex 4).

The staff payroll and other minimal direct costs associated with the PhD research (e.g. participation in conferences) were covered until 31st December 2024 by the Micronutrient Action Policy Support (MAPS) project. The MAPS project is funded by the Bill & Melinda Gates Foundation (INV-002855). From February 2022 to December 2022 and from February 2024 to date (2024), in addition to the MAPS funding, I was part-time seconded to the University of Nottingham.

References

Abdu, A. O., Kumssa, D. B., Joy, E. J. M., Groote, H. D., Lark, R. M., Broadley, M. R., & Gashu, D. (2022). Estimates of Dietary Mineral Micronutrient Supply from Staple Cereals in Ethiopia at a District Level. *Nutrients*, *14*(17), 3469. https://doi.org/10.3390/nu14173469

Arsenault, J. E., Moursi, M., Olney, D. K., Becquey, E., & Ganaba, R. (2020). Validation of 24-h dietary recall for estimating nutrient intakes and adequacy in adolescents in Burkina Faso. Maternal & Child Nutrition, 16(4), e13014. https://doi.org/10.1111/mcn.13014

Bailey, R. L., West Jr, K. P., & Black, R. E. (2015). The Epidemiology of Global Micronutrient Deficiencies. *Annals of Nutrition and Metabolism*, *66*(Suppl. 2), 22–33. https://doi.org/10.1159/000371618

Beal, T., Manohar, S., Miachon, L., & Fanzo, J. (2024). Nutrient-dense foods and diverse diets are important for ensuring adequate nutrition across the life course. *Proceedings of the National Academy of Sciences*, *121*(50), e2319007121. https://doi.org/10.1073/pnas.2319007121

Beal, T., Massiot, E., Arsenault, J. E., Smith, M. R., & Hijmans, R. J. (2017). Global trends in dietary micronutrient supplies and estimated prevalence of inadequate intakes. *PLOS ONE*, *12*(4), e0175554. https://doi.org/10.1371/journal.pone.0175554

Belay, A., Joy, E. J. M., Chagumaira, C., Zerfu, D., Ander, E. L., Young, S. D., Bailey, E. H., Lark, R. M., Broadley, M. R., & Gashu, D. (2020). Selenium Deficiency Is Widespread and Spatially Dependent in Ethiopia. *Nutrients*, *12*(6), Article 6. https://doi.org/10.3390/nu12061565

Bergau, S., Loos, T. K., & Sariyev, O. (2022). On- and Off-Farm Diversification and Travel Time to Markets: Linkages to Food Security in Rural Ethiopia. *The European Journal of Development Research*, *34*(5), 2543–2560. https://doi.org/10.1057/s41287-021-00475-0

Bhatnagar, K., Jha, K., Dalal, N., Patki, N., Gupta, G., Kumar, A., Kumar, A., & Chaudhary, S. (2024). Exploring micronutrients and microbiome synergy: Pioneering new paths in cancer therapy. *Frontiers in Immunology*, *15*. https://doi.org/10.3389/fimmu.2024.1442788

Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., de Onis, M., Ezzati, M., Mathers, C., & Rivera, J. (2008). Maternal and child undernutrition: Global and regional exposures and health consequences. *The Lancet*, *371*(9608), 243–260. https://doi.org/10.1016/S0140-6736(07)61690-0

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R., & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*, *382*(9890), 427–451. https://doi.org/10.1016/S0140-6736(13)60937-X

Botoman, L., Chagumaira, C., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Chimungu, J. G., Gameda, S., Gashu, D., Haefele, S. M., Joy, E. J. M., Kumssa, D. B., Ligowe, I. S., McGrath, S. P., Milne, A. E., Munthali, M., Towett, E., Walsh, M. G., Wilson, L., … Nalivata, P. C. (2022). Soil and landscape factors influence geospatial variation in maize grain zinc concentration in Malawi. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-12014-w

Broadley, M. R., White, P. J., Bryson, R. J., Meacham, M. C., Bowen, H. C., Johnson, S. E., Hawkesford, M. J., McGrath, S. P., Zhao, F.-J., Breward, N., Harriman, M., & Tucker, M. (2006). Biofortification of UK food crops with selenium. *Proceedings of the Nutrition Society*, *65*(2), 169–181. https://doi.org/10.1079/PNS2006490

Brown, K. H., Moore, S. E., Hess, S. Y., McDonald, C. M., Jones, K. S., Meadows, S. R., Manger, M. S., Coates, J., Alayon, S., & Osendarp, S. J. M. (2021). Increasing the availability and utilization of reliable data on population micronutrient (MN) status globally: The MN Data Generation Initiative. *The American Journal of Clinical Nutrition*, *nqab173*. https://doi.org/10.1093/ajcn/nqab173

Bruyn, J. de, Ferguson, E., Allman-Farinelli, M., Darnton-Hill, I., Maulaga, W., Msuya, J., Alders, R., de Bruyn, J., Ferguson, E., Allman-Farinelli, M., Darnton-Hill, I., Maulaga, W., Msuya, J., & Alders, R. (2016). Food composition tables in resource-poor settings: Exploring current limitations and opportunities, with a focus on animal-source foods in sub-Saharan Africa. *The British Journal of Nutrition*, *116*(10), 1709–1719. https://doi.org/10.1017/S0007114516003706

Charrondière, R. U., Stadlmayr, B., Rittenschober, D., Mouille, B., Nilsson, E., Medhammar, E., Olango, T., Eisenwagen, S., Persijn, D., Ebanks, K., Nowak, V., Du, J., & Burlingame, B. (2013). FAO/INFOODS food composition database for biodiversity. *Food Chemistry*, *140*(3), 408–412. https://doi.org/10.1016/j.foodchem.2012.08.049

Chilimba, A. D. C., Young, S. D., Black, C. R., Rogerson, K. B., Ander, E. L., Watts, M. J., Lammel, J., & Broadley, M. R. (2011). Maize grain and soil surveys reveal suboptimal dietary selenium intake is widespread in Malawi. *Scientific Reports*, *1*(1), 72. https://doi.org/10.1038/srep00072

Coates, J., Colaiezzi, B., Bell, W., Wafa, S., Prata, C., & Rogers, B. (2017). New dietary assessment technology for individual-level dietary data collection in low-income countries: The development process of inddex24. *Annals of Nutrition and Metabolism*, *71*(Supplement 2), 1129–1130. https://doi.org/10.1159/000480486

Coates, J., Colaiezzi, B., Fiedler, J. L., Wirth, J., Lividini, K., & Rogers, B. (2012). A Program Needs-Driven Approach to Selecting Dietary Assessment Methods for Decision-Making in Food Fortification

Programs. *Food and Nutrition Bulletin*, *33*(3_suppl2), S146–S156. https://doi.org/10.1177/15648265120333S202

Combs, G. F. (2015). Biomarkers of Selenium Status. *Nutrients*, *7*(4), Article 4. https://doi.org/10.3390/nu7042209

Dunlop, E., Cunningham, J., Adorno, P., Fatupaito, S., Johnson, S. K., & Black, L. J. (2025). Development of an updated, comprehensive food composition database for Australian-grown horticultural commodities. *Food Chemistry*, *463*, 141398. https://doi.org/10.1016/j.foodchem.2024.141398

Fairweather-Tait, S. J., Bao, Y., Broadley, M. R., Collings, R., Ford, D., Hesketh, J. E., & Hurst, R. (2011). Selenium in Human Health and Disease. *Antioxidants & Redox Signaling*, *14*(7), 1337–1383. https://doi.org/10.1089/ars.2010.3275

FAO & Government of Kenya. (2018). *Kenya Food Composition Tables*. http://www.fao.org/3/I9120EN/i9120en.pdf

FAO & Intake. (2022). *Global report on the state of dietary data*. FAO. https://doi.org/10.4060/cb8679en

FAO/INFOODS. (2012a). *FAO / INFOODS Guidelines for Food Matching Version 1.2*. https://www.fao.org/3/ap805e/ap805e.pdf

FAO/INFOODS. (2012b). *FAO/INFOODS Guidelines for Checking Food Composition Data prior to the Publication of a User Table/Database-Version 1.0*. FAO. https://www.fao.org/3/ap810e/ap810e.pdf

FAO & WFP. (2018). *Home-grown school feeding. Resource framework. Technical Document*. Food and Agriculture Organization of the United Nations: World Food Programme.

Ferguson, E. L., Gibson, R. S., Thompson, L. U., & Ounpuu, S. (1989). Dietary calcium, phytate, and zinc intakes and the calcium, phytate, and zinc molar ratios of the diets of a selected group of East African children. *The American Journal of Clinical Nutrition*, *50*(6), 1450–1456.

Galani, Y. J. H., Orfila, C., & Gong, Y. Y. (2020). A review of micronutrient deficiencies and analysis of maize contribution to nutrient requirements of women and children in Eastern and Southern Africa. *Critical Reviews in Food Science and Nutrition*, *0*(0), 1–24. https://doi.org/10.1080/10408398.2020.1844636

Gashu, D., Nalivata, P. C., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A. E., Mossa, A. W., Munthali, M., … Broadley, M. R. (2021). The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*, *594*(7861), Article 7861. https://doi.org/10.1038/s41586-021-03559-3

Gashu, D., Marquis, G. S., Bougma, K., & Stoecker, B. J. (2019). Spatial Variation of Human Selenium in Ethiopia. *Biological Trace Element Research*, *189*(2), 354–360. https://doi.org/10.1007/s12011-018-1489-5

Gibson, R. S. & Ferguson, E. (2008). *An interactive 24-hour recall for assessing the adequacy of iron and zinc intakes in developing countries*. 160.

Greenfield, H. & Southgate, D. A. T. (2003). *Food composition data: Production, management, and use*. FAO.

Government of Malawi. (2024). *National Agriculture Policy 2024*. https://agriculture.gov.mw/files/medias/national-agriculture-policy-2024-signed-1739192665.pdf

Hailu, H., Mamo, T., Keskinen, R., Karltun, E., Gebrekidan, H., & Bekele, T. (2015). Soil fertility status and wheat nutrient content in Vertisol cropping systems of central highlands of Ethiopia. *Agriculture & Food Security*, *4*(1), 19. https://doi.org/10.1186/s40066-015-0038-0

Hailu, K., Joy, E. J. M., Ferguson, E. L., Bailey, E. H., Wilson, L., Davis, K., Broadley, M. R., & Gashu, D. (2023). Dietary selenium intake among Ethiopian children in areas known for selenium spatial variability. *Frontiers in Nutrition*, *10*, 1250002. https://doi.org/10.3389/fnut.2023.1250002

Hicks, C. C., Cohen, P. J., Graham, N. A. J., Nash, K. L., Allison, E. H., D'Lima, C., Mills, D. J., Roscher, M., Thilsted, S. H., Thorne-Lyman, A. L., & MacNeil, M. A. (2019). Harnessing global fisheries to tackle micronutrient deficiencies. *Nature*, *574*(7776), Article 7776. https://doi.org/10.1038/s41586-019-1592-6

Hurst, R., Siyame, E. W. P., Young, S. D., Chilimba, A. D. C., Joy, E. J. M., Black, C. R., Ander, E. L., Watts, M. J., Chilima, B., Gondwe, J., Kang'ombe, D., Stein, A. J., Fairweather-Tait, S. J., Gibson, R. S., Kalimbira, A. A., & Broadley, M. R. (2013). Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi. *Scientific Reports*, *3*, 1425–1425. Africa-Wide Information. https://doi.org/10.1038/srep01425

Hutchinson, J., Rippin, H., Threapleton, D., Jewell, J., Kanamäe, H., Salupuu, K., Caroli, M., Antignani, A., Pace, L., Vassallo, C., Lande, B., Hildonen, C., Rito, A. I., Santos, M., Gabrijelcic Blenkus, M., Sarkadi-Nagy, E., Erdei, G., Cade, J. E., & Breda, J. (2021). High sugar content of European commercial baby foods and proposed updates to existing recommendations. *Maternal & Child Nutrition*, *17*(1), e13020. https://doi.org/10.1111/mcn.13020

Ingenbleek, L., Jazet, E., Dzossa, A. D., Adebayo, S. B., Ogungbangbe, J., Dansou, S., Diallo, Z. J., Kouebou, C., Adegboye, A., Hossou, E., Coulibaly, S., Eyangoh, S., Le Bizec, B., Verger, P., Kamanzi, J., Merten, C., & Leblanc, J.-C. (2017). Methodology design of the regional Sub-Saharan Africa Total Diet Study in Benin, Cameroon, Mali and Nigeria. *Food and Chemical Toxicology*, *109*(Pt 1), 155–169. Africa-Wide Information.

Joy, E. J. M., Ander, E. L., Young, S. D., Black, C. R., Watts, M. J., Chilimba, A. D. C., Chilima, B., Siyame, E. W. P., Kalimbira, A. A., Hurst, R., Fairweather-Tait, S. J., Stein, A. J., Gibson, R. S., White, P. J., & Broadley, M. R. (2014). Dietary mineral supplies in Africa. *Physiologia Plantarum*, *151*(3), 208–229.

Joy, E. J. M., Broadley, M. R., Young, S. D., Black, C. R., Chilimba, A. D. C., Ander, E. L., Barlow, T. S., & Watts, M. J. (2015). Soil type influences crop mineral composition in Malawi. *Science of the Total Environment*, *505*, 587–595. https://doi.org/10.1016/j.scitotenv.2014.10.038

Joy, E. J. M., Kumssa, D. B., Broadley, M. R., Watts, M. J., Young, S. D., Chilimba, A. D. C., & Ander, E. L. (2015). Dietary mineral supplies in Malawi: Spatial and socioeconomic assessment. *BMC Nutrition*, *1*(1), 42. https://doi.org/10.1186/s40795-015-0036-4

Kihara, J., Bolo, P., Kinyua, M., Rurinda, J., & Piikki, K. (2020). Micronutrient deficiencies in African soils and the human nutritional nexus: Opportunities with staple crops. *Environmental Geochemistry and Health*, *42*(9), 3015–3033. https://doi.org/10.1007/s10653-019-00499-w

Kumssa, D. B., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Chimungu, J. G., Davis, K., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A., Muleya, P., Munthali, M., … Nalivata, P. C. (2022). Cereal grain mineral micronutrient and soil chemistry data from GeoNutrition surveys in ethiopia and malawi. *Scientific Data*, *9*(1), Article 1. https://doi.org/10.1038/s41597-022-01500-5

Leclercq, C., Allemand, P., Balcerzak, A., Branca, F., Ferreira de Sousa, R., Lartey, A., Lipp, M., Padula de Quadros, V., & Verger, P. (2019). FAO/WHO GIFT (Global Individual Food consumption data Tool): A global repository for harmonised individual quantitative food consumption studies. *Proceedings of the Nutrition Society*, *78*, 1–12. https://doi.org/10.1017/S0029665119000491

Lephole, M., Khaketla, M., & Monoto, M. (2006). *Lesotho Food Composition Table*. Maseru: Department of Agricultural Research.

Ligowe, I. S., Phiri, F. P., Ander, E. L., Bailey, E. H., Chilimba, A. D. C., Gashu, D., Joy, E. J. M., Lark, R. M., Kabambe, V., Kalimbira, A. A., Kumssa, D. B., Nalivata, P. C., Young, S. D., & Broadley, M. R. (2020). Selenium deficiency risks in sub-Saharan African food systems and their geospatial linkages. *Proceedings of the Nutrition Society*, *79*(4), 457–467. https://doi.org/10.1017/S0029665120006904

Ma, X.-M., Li, K.-X., Chen, Z.-Q., Wu, C.-M., Liao, W.-Z., & Guo, X.-G. (2024). Impact of age, sex, and thyroid autoimmunity on the association between selenium intake and type 2 diabetes mellitus. *BMC Public Health*, *24*(1), 743. https://doi.org/10.1186/s12889-024-18225-2

Micha, R., Coates, J., Leclercq, C., Charrondière, U. R., & Mozaffarian, D. (2018). Global Dietary Surveillance: Data Gaps and Challenges. *Food and Nutrition Bulletin*, *39*(2), 175–205. https://doi.org/10.1177/0379572117752986

Muleya, M., Young, S. D., Reina, S. V., Ligowe, I. S., Broadley, M. R., Joy, E. J. M., Chopera, P., & Bailey, E. H. (2021). Selenium speciation and bioaccessibility in Se-fertilised crops of dietary importance in Malawi. *Journal of Food Composition and Analysis*, *98*, 103841. https://doi.org/10.1016/j.jfca.2021.103841

Naderi, M., Puar, P., Zonouzi-Marand, M., Chivers, D. P., Niyogi, S., & Kwong, R. W. M. (2021). A comprehensive review on the neuropathophysiology of selenium. *Science of The Total Environment*, *767*, 144329. https://doi.org/10.1016/j.scitotenv.2020.144329

National Statistical Office (NSO). (2017). *Fourth Integrated Household Survey 2016-2017* [Data set]. World Bank, Development Data Group. https://doi.org/10.48529/G2P9-9R19

National Statistical Office (NSO) [Malawi], & ICF. (2017). *Malawi Demographic and Health Survey 2015-16*. https://dhsprogram.com/pubs/pdf/FR319/FR319.pdf

National Statistical Office and ICF. (2016a). Malawi Demographic and Health Survey 2016 [Dataset]. Individual level MWIR7AFL.dta. Rockville, Maryland: General Directorate of Statistics, Ministry of Finance, and ICF [Producers]. ICF [Distributor], 2013.

National Statistical Office and ICF. (2016b). Malawi Demographic and Health Survey 2016 [Dataset]. Biomarkers. MW_WRA.dta. Rockville, Maryland: General Directorate of Statistics, Ministry of Finance, and ICF [Producers]. ICF [Distributor], 2013.

Neufeld, L. M., Baker, S., Garrett, G. S., & Haddad, L. (2017). Coverage and Utilization in Food Fortification Programs: Critical and Neglected Areas of Evaluation. *The Journal of Nutrition*, *147*(5), 1015S-1019S. https://doi.org/10.3945/jn.116.246157

Ngigi, P. B., Du Laing, G., Masinde, P. W., & Lachat, C. (2020). Selenium deficiency risk in central Kenya highlands: An assessment from the soil to the body. *Environmental Geochemistry and Health*, *42*(7), 2233–2250. https://doi.org/10.1007/s10653-019-00494-1

Nguyen, T.-T., & Qaim, M. (2025). Local and regional food production diversity are positively associated with household dietary diversity in rural Africa. *Nature Food*, *6*(2), 205–212. https://doi.org/10.1038/s43016-024-01096-6

Noisel, N., Carrier, G., & Bouchard, M. (2014). Study of selenium intake and disposition in various matrices based on mathematical algorithms derived from pooled biomonitoring data. *International Journal of Hygiene and Environmental Health*, *217*(7), 796–804. https://doi.org/10.1016/j.ijheh.2014.04.005

Phiri, F. P., Ander, E. L., Bailey, E. H., Chilima, B., Chilimba, A. D. C., Gondwe, J., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Phuka, J. C., Salter, A., Suchdev, P. S., Watts, M. J., Young, S. D., & Broadley, M. R. (2019). The risk of selenium deficiency in Malawi is large and varies over multiple spatial scales. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-43013-z

Phiri, F. P., Ander, E. L., Lark, R. M., Bailey, E. H., Chilima, B., Gondwe, J., Joy, E. J. M., Kalimbira, A. A., Phuka, J. C., Suchdev, P. S., Middleton, D. R. S., Hamilton, E. M., Watts, M. J., Young, S. D., & Broadley, M. R. (2020). Urine selenium concentration is a useful biomarker for assessing population level selenium status. *Environment International*, *134*, 105218. https://doi.org/10.1016/j.envint.2019.105218

Public Health England. (2021). McCance and Widdowson's composition of foods integrated dataset.

Rayman, M. P. (2020). Selenium intake, status, and health: A complex relationship. *Hormones*, *19*(1), 9–14. https://doi.org/10.1007/s42000-019-00125-5

Stefanowicz, F. A., Talwar, D., O'Reilly, D. S. J., Dickinson, N., Atkinson, J., Hursthouse, A. S., Rankin, J., & Duncan, A. (2013). Erythrocyte selenium concentration as a marker of selenium status. *Clinical Nutrition*, *32*(5), 837–842. https://doi.org/10.1016/j.clnu.2013.01.005

Stevens, G. A., Beal, T., Mbuya, M. N. N., Luo, H., Neufeld, L. M., Addo, O. Y., Adu-Afarwuah, S., Alayón, S., Bhutta, Z., Brown, K. H., Jefferds, M. E., Engle-Stone, R., Fawzi, W., Hess, S. Y., Johnston, R., Katz, J., Krasevec, J., McDonald, C. M., Mei, Z., … Young, M. F. (2022). Micronutrient deficiencies among preschool-aged children and women of reproductive age worldwide: A pooled analysis of individual-level data from population-representative surveys. *The Lancet Global Health*, *10*(11), e1590–e1599. https://doi.org/10.1016/S2214-109X(22)00367-9

Tang, K., Adams, K. P., Ferguson, E. L., Woldt, M., Kalimbira, A. A., Likoswe, B., Yourkavitch, J., Chrisinger, B., Pedersen, S., Segovia De La Revilla, L., Dary, O., Ander, E. L., & Joy, E. J. M. (2021). Modeling food fortification contributions to micronutrient requirements in Malawi using Household Consumption and Expenditure Surveys. *Annals of the New York Academy of Sciences*, *n/a*(n/a). https://doi.org/10.1111/nyas.14697

Tang, K., Adams, K. P., Ferguson, E. L., Woldt, M., Yourkavitch, J., Pedersen, S., Broadley, M. R., Dary, O., Ander, E. L., & Joy, E. J. M. (2022). Systematic review of metrics used to characterize dietary nutrient supply from household consumption and expenditure surveys. *Public Health Nutrition*, 1–32. https://doi.org/10.1017/S1368980022000118

Thomson, C. D. (2004). Assessment of requirements for selenium and adequacy of selenium status: A review. *European Journal of Clinical Nutrition*, *58*(3), 391–402. https://doi.org/10.1038/sj.ejcn.1601800

Traka, M. H., Plumb, J., Berry, R., Pinchen, H., & Finglas, P. M. (2020). Maintaining and updating food composition datasets for multiple users and novel technologies: Current challenges from a UK perspective. *Nutrition Bulletin*, *45*(2), 230–240. https://doi.org/10.1111/nbu.12433

United States Department of Agriculture (USDA), 2019. Agricultural Research Service, Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Legacy. Version Current: April 2019. Washington, DC, Agricultural Research Service, Nutrient Data Laboratory. https://fdc.nal.usda.gov/ download-datasets.html

van Graan, A., Joelaine, C., Malory, J., Sitilitha, M., Agnes, M., Felix P., P., Lynne M., A., Shibani, G., & Elizabeth, M.-C. (2019). *Malawian Food Composition Tables*.

Vincent, A., Grande, F., Compaoré, E.,U.R., Amponsah Annor, G., Addy, P.A., Aburime, L.C., Ahmed, D., Bih Loh, A.M, Dahdouh Cabia, S., Deflache, N., Dembélé, F.M., Dieudonné, B., Edwige, O.B., Ene-Obong, H.N, Fanou Fogny, N., Ferreira, M., Omaghomi Jemide, J., Kouebou, P.C., Muller, C., … Charrondière, U.R. (2020). *FAO/INFOODS Food Composition Table for Western Africa (2019) User Guide & Condensed Food Composition Table*.

Watts, M. J., Middleton, D. R. S., Marriott, A. L., Humphrey, O. S., Hamilton, E. M., Gardner, A., Smith, M., McCormack, V. A., Menya, D., Munishi, M. O., Mmbaga, B. T., & Osano, O. (2019). Source apportionment of micronutrients in the diets of Kilimanjaro,Tanzania and Counties of Western Kenya. *Scientific Reports*, *9*(1), 14447–14447. https://doi.org/10.1038/s41598-019-51075-2

World Health Organization. (2021). *Report of the technical consultation on measuring healthy diets: Concepts, methods and metrics: virtual meeting, 18–20 May 2021*. World Health Organization. https://apps.who.int/iris/handle/10665/351048

3 Chapter 3: Assessing the availability and geographic location of food composition data used to estimate micronutrient intakes in sub-Saharan Africa: A scoping review

This chapter presents the scoping review which evaluated the availability of mineral micronutrient composition data and accompanying metadata, including the geographic location of the samples for use in sub-Saharan Africa.

It was published in the *Journal of Food Composition of Analysis* and can be accessed here (https://doi.org/10.1016/j.jfca.2023.105322). In addition, following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines a protocol was developed and published as a preprint in the *Open Science Framework* (OSF) and can be accessed here (https://doi.org/10.31219/osf.io/vd2mf).

The repository to replicate the analysis is publicly available here (https://doi.org/10.5281/zenodo.10213390).

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	2006000	**Title**	Ms
First Name(s)	Lucia		
Surname/Family Name	Segovia de la Revilla		
Thesis Title	Geospatial analysis of food composition data to estimate population micronutrient intake		
Primary Supervisor	Edward Joy		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Journal of Food Composition and Analysis		
When was the work published?	6 April 2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conceptualised, performed the analysis and was the primary writer. Edward Joy, Elaine Ferguson and Claire Dooley helped with the conceptualisation, supervision and revision of the manuscript. Gareth Osman helped with the revision of the search and code, and Louise Ander helped reviewing the final draft of the manuscript.

SECTION E

Student Signature	Lucia Segovia de la Revilla
Date	27/10/2024

Supervisor Signature	EJMJ - Edward Joy
Date	06/01/2025

Evidence of retention of copyright

"An open science framework and tools to create reproducible food composition data for use in nutrition"

In line with the Elsevier OA with CC-BY, I have the right to re-use the article without permission from Elsevier, see the Copyright page for more information.

Authors' rights in the article

	OA with CC BY	OA with CC BY-NC	OA with CC BY-NC-ND	Subscription
Receive proper attribution and credit for their published work	Yes	Yes	Yes	Yes
Re-use their article in their own new works, without permission from Elsevier or payment to Elsevier, including by: • making copies of the article (or part of the article) to promote companies or products they own, whether or not such promotion is commercial; • including the article in a thesis or dissertation; • extending the article to a book, including the article in a subsequent compilation of their own work, or re-using portions, excerpts, and their own figures, tables and images from the article in their own new works (which in each case may be published with Elsevier or with a third party commercial or non-commercial publisher, at the author's discretion)	Yes	Yes	Yes	Yes[1]

Journal of Food Composition and Analysis 120 (2023) 105322

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

The availability and geographic location of open-source food composition data used to estimate micronutrient intakes in sub-Saharan Africa: A scoping review

Lucia Segovia de la Revilla [a], Elaine L. Ferguson [a], Claire Dooley [a], Gareth Osman [b], E. Louise Ander [c,d], Edward J.M. Joy [a,e,*]

[a] *Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom*
[b] *Department of Human Nutrition and Health, Lilongwe University of Agriculture and Natural Resources, Malawi*
[c] *School of Biosciences, University of Nottingham. Bonington Hall, Sutton Bonington, Loughborough LE12 5RD, United Kingdom*
[d] *Centre for Environmental Geochemistry, British Geological Survey, Keyworth, Nottingham NG12 5GG, United Kingdom*
[e] *Sustainable Soils and Crops Department, Rothamsted Research, West Common, Harpenden AL5 2JQ, United Kingdom*

A R T I C L E I N F O

Keywords:
Food composition data
INFOODS
Metadata
Minerals
Micronutrients
Spatial variability
Sub-Saharan Africa

A B S T R A C T

Background: Estimates of dietary micronutrient intakes rely on food composition data. The nutrient composition of foods varies spatially with potentially large effects on dietary micronutrient intakes. This review assessed the availability and geographic origin of five minerals (calcium, iron, iodine, selenium and zinc) in publicly available food composition tables/databases (FCTs) for use in sub-Saharan Africa (SSA).
Methods: A scoping review was conducted following PRISMA guidelines, in which four databases (MEDLINE, Embase, Global Health and Africa Wide Information) and four online resources were searched to identify published FCTs for use in SSA. Metadata were reviewed to identify the geographic origin of composition values for selected foods.
Results: Nineteen publicly available FCTs were identified, with the highest geographic coverage in Eastern Africa (45% of countries) and lowest coverage in Central Africa (12% of countries). Iodine and selenium were reported in four and six FCTs, respectively, while iron and calcium were included in $\geq$ 18 FCTs. More than 60% of nutrient values were borrowed from other FCTs. The geographic origin of 22% of mineral values were documented.
Conclusions: Limited local food composition analytical data is available, for estimating mineral intakes of SSA populations, with poor documentation of the data sources and the geographic origins of samples. New data structures and improved metadata are required to capture and report geographic information in publicly available FCTs, and to accommodate a new generation of spatially-resolved food composition data.

1. Introduction

Globally, over two billion people are affected by one or more micronutrient deficiencies (MNDs). The prevalence of MNDs and associated disease burdens is highest in low- and middle-income countries, including sub-Saharan Africa (Arsenault et al., 2015; Beal et al., 2017; Joy et al., 2014; Schmidhuber et al., 2018; Smith et al., 2016; Wessells et al., 2012; White et al., 2021). Inadequate dietary micronutrient intake is one of the factors contributing to the high prevalence of MNDs. Hence, reliable data on the micronutrient composition of foods, food consumption patterns, dietary intakes of micronutrients and when available, the prevalence of MNDs are required for designing, prioritising, monitoring, and evaluating programmes and policies to alleviate micronutrient deficiencies, including biofortification and food fortification (Neufeld et al., 2017; Popkin et al., 2020).

In recent years, several groups have compiled regional and national food composition tables and databases (FCTs) and nutrient conversion tables (NCTs) (see Table 1: Food composition terms and definitions) to assess the dietary micronutrient intakes of populations in sub-Saharan Africa (e.g., FAO/INFOODS Food Composition Table for Western

Abbreviations: FCTs, Food Composition Tables and Databases; NCTs, Nutrient Conversion Tables; MNDs, Micronutrient Deficiencies.

* Correspondence to: Department of Population Health Faculty of Epidemiology & Population Health London School of Hygiene & Tropical Medicine, Keppel St, London WC1E 7HT, United Kingdom.
E-mail address: Edward.joy@lshtm.ac.uk (E.J.M. Joy).

https://doi.org/10.1016/j.jfca.2023.105322
Received 23 January 2023; Received in revised form 3 April 2023; Accepted 4 April 2023
Available online 6 April 2023

Table 1
Food composition terms and definitions.

Term	Definition	Reference
Food composition tables and databases (FCTs)	FCTs are lists of foods, in (printed) tables and/or database formats, that provide data on the content of energy, nutrients and other food components. Ideally the selection of foods included would cover those foods that are relevant and highly consumed by the population for which the FCTs are intended for.	Adapted from Greenfield and Southgate (2003).
Type of (food composition) data	FCTs are normally compiled using combined methods. Some values may be obtained form direct chemical analysis of foods (primary data), others by compilation from literature screening (secondary data), or borrowing from other FCTs	Adapted from Greenfield and Southgate (2003).
Primary analytical data	This type of value is the result of chemical analysis of food carried out specifically to populate a FCT.	Adapted from Greenfield and Southgate (2003).
Secondary analytical data	This type of value is taken from published literature, thesis, unpublished laboratory reports etc. that reported chemical analysis of foods.	Adapted from Greenfield and Southgate (2003).
Borrowed data	This type of value is taken from other FCTs.	Adapted from Greenfield and Southgate (2003).
Source of the data	It refers to where the data were obtained from (e.g., a study published in literature, or contractors' analysis).	Adapted from Greenfield and Southgate (2003) and Pennington (2008) .
Nutrient Conversion Tables (NCTs)	A nutrient conversion Table (NCT) is a collection of data on the nutrient content of foods reported as supplied, consumed and/or acquired in a specific survey, e.g. a household consumption and expenditure survey. Therefore, the NCT is study specific, and it is based on compiled information from national and/or regional food composition tables and databases (FCTs/FCDBs) following food matching guidelines (e.g., FAO/INFOODS). It is needed to assess the dietary energy and nutrient supply and/or (apparent) consumption of the survey's population.	Developed for this study, based on Moltedo et al. (2018), and personal communication with the author (A.M).

Africa (2019), Malawian Food Composition Table (2019), Kenya Food Composition Tables (2018), etc.). While considerable efforts have been made to compile high quality and comprehensive food composition data, limitations remain. For example, nutrient values are typically borrowed from the FCTs of other countries when analysed data from in-country food samples are not available. These borrowed values may be inaccurate due to spatial variation in crop composition, as well as differences in food fortification policies and local food preparation and cooking practices (Ene-Obong et al., 2019; van Graan et al., 2019; Pennington et al., 2007). Furthermore, the nutrient composition values of each food item, (e.g., iron in wheat flour, or zinc in maize flour), are typically represented by a single data point, which may mask important subnational variation, including spatially-structured variation in the nutrient content of local foods (Gashu et al., 2021). Additionally, re-use of data may propagate nutrient composition values which were quantified using obsolete analytical methods and equipment, or values with inadequate descriptions and/or reporting of nutrients, foods and recipes (Traka et al., 2020). Finally, in some FCTs and for some nutrients, there can be multiple missing values. For example, in the Food Composition Table for use in Africa, which is still used as a source of mineral values, only 50% of the 1624 food items have data on mineral composition (Woot-Tsuen et al., 1968). This paucity of data and the lack of transparency in nutrient composition metadata has major implications for the ability to accurately estimate the prevalence of the population at risk of inadequate dietary micronutrient intakes.

The nutrient composition of crops vary spatially due to factors including soil type and climate, and this variation can be nutritionally relevant. For some minerals, spatial variation in staple crop composition is one of the main drivers of sub-national variation in population status (Belay et al., 2020; Botoman et al., 2022; Gashu et al., 2021; Kihara et al., 2020; Phiri et al., 2019). This is particularly true in contexts where food systems are localised and diets are dominated by locally produced staple crops (e.g., maize, rice, *teff,* cassava), which is the case for many people living in sub-Saharan African countries (Joy et al., 2015b; Ryckman et al., 2021). When foods are produced and sourced locally, their nutrient content will partly depend on local agri-food practices (e.g., use of inputs, milling, fermentation, storage, etc.) and environmental factors (e.g., soil type, soil pH, precipitation, etc.). Several studies in sub-Saharan Africa have reported spatial variability in the mineral contents of various staple crops. For example, Gashu and colleagues (2021) reported nutritionally important geospatial variation in the content of calcium (Ca), iron (Fe), selenium (Se), and zinc (Zn) in *teff* and wheat grains in Ethiopia and in maize grain in Malawi. Smaller studies have investigated the spatial variation in other countries in the region, yielding similar conclusions (Joy et al., 2015a;b; Manzeke et al., 2019; Watts et al., 2019; Wood et al., 2018). These studies highlight the importance of locally-sampled food composition data, including the five minerals included in this review: Ca, Fe, iodine (I), Se and Zn, for which spatial variation in their concentration may influence the estimates of inadequate and/or excessive intakes resulting in important health implications. Close monitoring of dietary mineral intakes (including from drinking water) and fortification programs in the region is vital to identify populations at risk of inadequate and/or excessive intakes. Thus, to accurately quantify and understand the prevalence and aetiology of mineral deficiencies, spatially structured food composition data are needed to estimate the intake of minerals including Ca, Fe, I, Se and Zn. The construction of these FCTs/NCTs should reflect local food systems and spatial variability at scales relevant to the context (Botoman et al., 2022; Gashu et al., 2021; Joy et al., 2015a,b; Ligowe et al., 2020).

Information on the type of data populating FCTs/NCTs and their geographic origin are necessary to assess the influence of spatial variation on mineral intake estimations for populations consuming locally produced foods. That information is also essential to evaluate modelling outputs that use FCTs/NCTs as underlying data, however it is currently unavailable and/or difficult to access. Hence, this scoping review aimed to capture the current status of freely available FCTs/NCTs used or available for use in estimating dietary intakes of Ca, Fe, I, Se and Zn in sub-Saharan Africa, and to identify data gaps and limitations. Specifically, the data and metadata available to assess spatial variation in the mineral content of foods were evaluated.

2. Materials and methods

2.1. Research questions

The objective of this review was to assess the availability of open-sourced and geographically-relevant FCTs/ NCTs to estimate dietary intakes of Ca, Fe, I, Se, and Zn in sub-Saharan Africa. These minerals were selected for their public health relevance given the widespread prevalence of inadequate dietary supplies in sub-Saharan Africa (Joy et al., 2014). There is also evidence of strong spatial variation in the mineral contents of staple foods, which may result in spatial variation in dietary adequacy given the localised nature of food systems, particularly for rural, low-income households (Gashu et al., 2021).

The primary research question was: What FCTs/ NCTs are freely available in the public domain for use in estimating dietary intakes of Ca, Fe, I, Se, and Zn in sub-Saharan Africa?

The secondary research questions were:

a. What is the geographic scope of the FCTs/ NCTs available to estimate dietary intakes of Ca, Fe, I, Se and Zn in sub-Saharan Africa?
b. What percentage of mineral composition data are missing for selected food items in FCTs/ NCTs available for use in sub-Saharan Africa?
c. For available data, what information is reported on the geographic location of food samples, and is it reported only at the food item level (e.g. maize flour) or at food item-micronutrient level (e.g. Zn in maize flour)?

2.2. Study design and protocol

A study protocol was developed following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for protocols (Moher et al., 2015). A full description of the method can be found in the study protocol which was registered with the Open Science Framework (OSF) (Segovia de la Revilla et al., 2022), hence only a brief description is provided below.

A scoping review method was used, which was informed by the PRISMA extension for scoping reviews (PRISMA-ScR) as guidelines (Tricco et al., 2018). FCTs and NCTs that provide data on Ca, Fe, I, Se and Zn concentration in foods for any geographic location within sub-Saharan Africa were identified. Mineral values, data documentation and metadata available were extracted for a selection of food items in each FCT/ NCT, including information on the geographic location of each mineral value per food item. The food items selected were those that matched the Food and Agriculture Organization of the United Nations (FAO) Food Balance Sheet (FBS) ([dataset] FAO, 2010) food categories providing a high percentage of energy and/or minerals in the food supply of sub-Saharan Africa countries, as described in detail in the study protocol (Section 4.1: "Food items included for the secondary outcomes" in Segovia de la Revilla et al., 2022) and summarised below in the Section 2.6.

2.3. Eligibility criteria

Stand-alone FCTs developed for sub-Saharan African countries and/ or regions were included, regardless of whether they had previously been used in a published study. In addition, studies that reported the use of FCTs/ NCTs to estimate dietary intakes of at least one of the minerals of interest (Ca, Fe, I, Se and Zn) by human subjects (with no age or gender restriction) in sub-Saharan Africa, including both experimental and non-experimental study designs (e.g., cohort, case-control, cross-sectional, ecological, modelling), were reviewed. From these studies, the sources of food composition data used in the study were identified and the relevant FCTs/ NCTs were retrieved. When multiple versions of a specific FCT were available, only the most recent version was reviewed, on the assumption that an update is based on the most recent and accurate information. Moreover, only those that were free to access and in digital format were included.

2.4. Information sources and search strategies

In this review four online databases were searched: MEDLINE, Embase, Global Health and Africa Wide Information. Keywords and subject headings, when available, were used to cover all terms related to food and nutritional composition data. In addition, "Expert Search" (i.e., "ALL countries in sub-Saharan Africa, Medline. List from World Bank, June 2019", "ALL sub-Saharan Africa Countries, Embase. List from World Bank, June 2019") were used to ensure that all countries in sub-Saharan Africa were included. No limits, search filters or restrictions were applied in any of the databases. In addition, four online resources were included in the search, which were: FAO/INFOODS: Africa (INFOODS, 2022), LanguaL (LanguaL, 2022), Nutritool (Nutritools, 2018), World Nutrient Databases for Dietary Studies (WNDDS) (WNDDS, 2022). The search strategy was designed following PRISMA guidelines (Moher et al., 2015), and the PRISMA extension for Reporting Literature Search in Systematic Reviews (PRISMA-S) (Rethlefsen et al., 2021). It was published together with the protocol in OSF (Segovia de la Revilla et al., 2022).

2.5. Study selection process

Screening and removal of duplicates was carried out using Mendeley (v.1.19.8). The screening of the studies was done in two steps: first, titles and abstracts were screened and second, from those selected in step one and retrieved, full text studies were screened. The reason(s) for exclusion were recorded in a spreadsheet (R1: excluded due to study type, R2: excluded due to nutrient(s) reported, R3: excluded due to location/scope of the study). Similarly, information on the retrieval of full text studies ("Yes"/"No") and the reason for failure (R1: not electronically available, R2: not found, R3: not accessible) were documented. Then, the food composition information was extracted from each study which included: FCT/ NCT name, authors, geographic location, and date of publication. Similar steps were carried out for stand-alone FCTs extracted from grey literature and websites. For FCTs that were not available online, a request was sent to the authors. Details on the data and the screening process are reported in the supplementary materials.

2.6. Data items and data abstraction process

The primary outcome was a list of the most recent, free to access and available FCTs/ NCTs for use in sub-Saharan Africa that reported at least one of the five minerals of interest (Ca, Fe, I, Se, and Zn), and the geographic scope of those datasets.

The secondary outcomes were the availability of mineral nutrient values, the source of information, type of data (See "Terms and definitions" in Table 1) and the geographic origin of each mineral value for a selection of foods. Data extracted to inform the secondary outcomes included the pre-defined subset of food items per FCT/ NCT. This subset of food items was selected based on FBS categories (e.g., maize and products, rice and products) and their contribution to the energy and/or the supply of the five minerals of interest in sub-Saharan Africa, as described elsewhere (Segovia de la Revilla, et al., 2022).

2.7. Synthesis of results

The list of all the FCTs/ NCTs for use in sub-Saharan Africa were collated in a spreadsheet. Similarly, the list of food items from each FCT/ NCT, their mineral content per 100 g and documentation, including type and source of data and geographic location, were extracted. Both data collection forms were piloted by extracting the data from three FCTs. More details regarding the data extraction and variables extracted are reported in Section 4 of the protocol (Segovia de la Revilla et al., 2022).

Data processing (except screening of the studies and de-duplication of the results), which included: compilation of the results from the screening process (i.e., FCTs/NCTs found in databases and in online databases and in resources), loading the data (FCTs/NCTs), extraction of the mineral values and the source of information per food item, harmonisation of the variable names, summary statistics (i.e., counts, mean, median, minimum and maximum) and visualisation (i.e., boxplots, histograms) were performed using R (R Core Team, 2020) and RStudio (Rstudio Team, 2020) and the code is available in GitHub.

3. Results

3.1. Study flow

The screening process is summarised in the PRISMA 2020 flow diagram (Fig. 1), and screening records can be found in supplementary materials (S.M.1). In summary, from the four online databases searched, 921 studies were identified for screening, of which 243 studies were selected for full text screening. Of these, 70 studies were unavailable, and 173 were reviewed for eligibility. Ultimately, from the electronic databases, 109 studies were included in the review and from those studies, 50 FCTs/ NCTs were identified that had been used to estimate mineral intakes of populations in sub-Saharan Africa.

From the four online resources (i.e., websites that hosted FCTs or FCT listings), 44 unique stand-alone FCTs were identified for sub-Saharan Africa (Fig. 1). No FCTs were retrieved from WNDDS as data were not accessible (that is, the underlying data displayed in the dashboard were unavailable) and all the FCTs displayed in the dashboard had been included.

After removal of duplicates, 71 unique FCTs/ NCTs were identified for use in estimating dietary intakes of Ca, Fe, I, Se and Zn in sub-Saharan Africa. From these, 44 were excluded based on the eligibility criteria of selecting the most recent version for each geographic area (i.e., country, region), while eight were excluded based on accessibility (Suppl. Tables 1–2).

3.2. Food composition data for use in sub-Saharan Africa

In total, 19 FCTs/ NCTs for use in sub-Saharan Africa were identified based on our criteria which dated from 1988 to 2021 (Table 2). In Fig. 2, the geographic scope of the FCTs/NCTs included in the review are presented. The majority of the FCTs/ NCTs included were national FCTs/ NCTs (n = 13), of which the highest coverage was provided for Eastern Africa (n = 7; 37% of 19 Eastern African countries), followed by Western Africa (n = 4; 23% of 17 Western African countries). For Southern and Central regions only one FCT and one NCT, in each region respectively, were included in this review (representing 20% of five Southern African and 10% of 10 Central African countries, respectively).

The widest geographic area covered by the FCTs/NCTs reviewed was the regional Africa NCT (Joy et al., 2014) which provided food composition data on the five minerals for three sub-Saharan African regions (Eastern, Western, and Southern regions), whereas the smallest geographic areas covered were the two sub-national NCTs for Kenya, Tanzania (Watts et al., 2019a,b) and two sub-national FCTs for Uganda

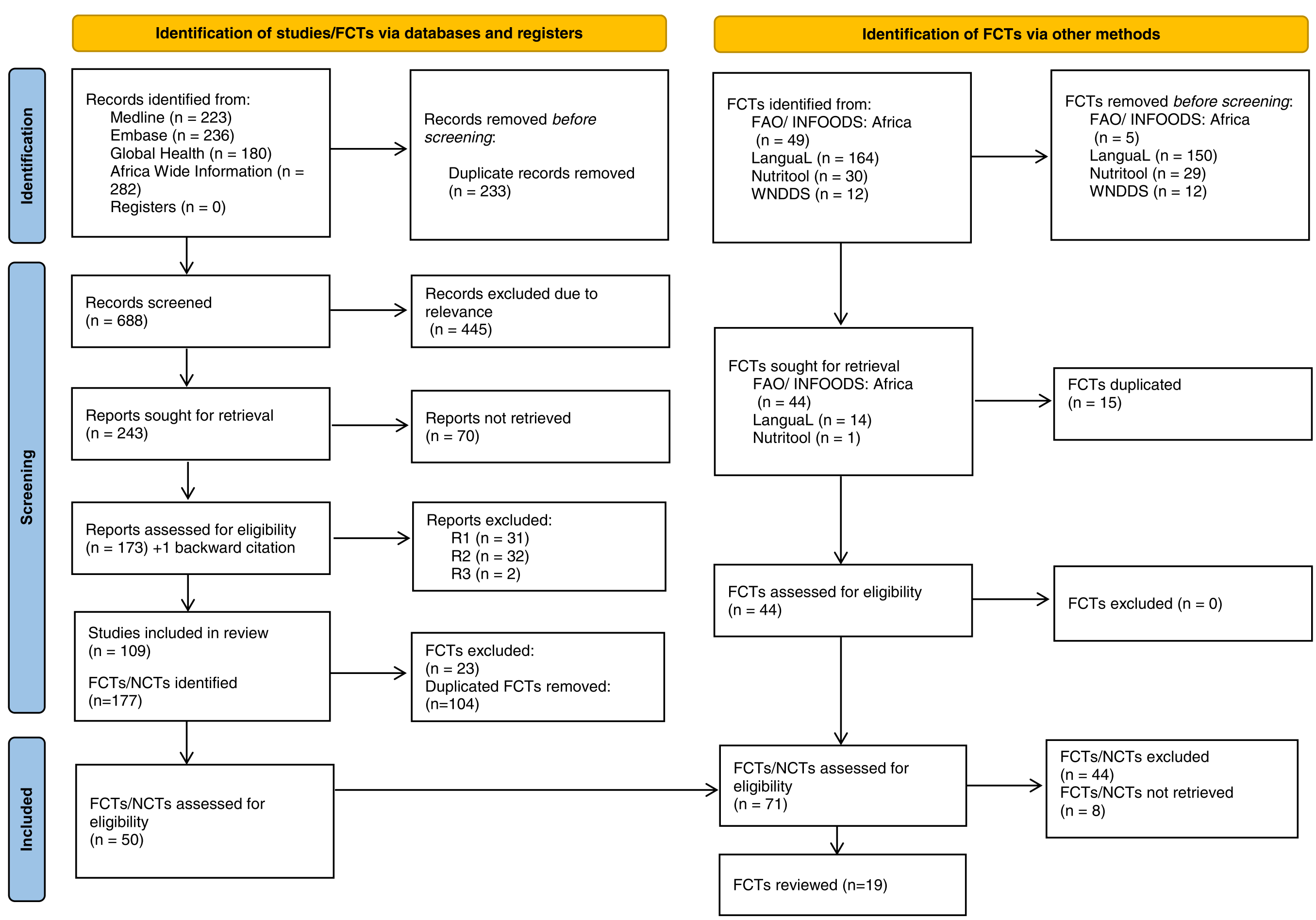


Fig. 1. *FCTs: Food Composition Tables and Databases, NCTs: Nutrient Conversion Tables, WNDDS: World Nutrient Databases for Dietary Studies.* Footnote: R1: excluded due to study type (e.g., not using or reporting FCT for dietary assessment), R2: excluded due to nutrient(s) reported, R3: excluded due to location/scope of the study. PRISMA – Flow diagram adapted from Page et al., 2021.

Table 2
List of food composition tables and databases (FCTs) and nutrient conversion tables (NCTs) included in the review and their characteristics.

Food composition table name	*Year of publication*	*Food Composition Type*	*Food Composition Format*	*Type of data*	*Language*	*Geographic Scope*	*Mineral reported*	*Reporting level*	*Food items reviewed*	*Reference*
Eastern Africa FCT, 1988	1988	Table	pdf	Multiple source (analytical, borrowed, calculated)	English	East Africa	Ca, Fe,	Mineral	18% (26/146)	West et al. (1988)
Zimbabwe FCT, 1989	1989	Table	pdf	Borrowed from FCTs, analytical values from literature	English	Zimbabwe	Ca, Fe,	Food Composition	13% (17/135)	Chitsiku (1989)
Lesotho FCT, 2006	2006	Table/database	pdf/xlsx	Borrowed from FCTs, analytical values from literature	English	Lesotho	Ca, Fe, Se, Zn	Item	6% (17/294)	Lephole et al. (2006)
Tanzania FCT, 2008	2008	Table/database	pdf/xlsx	Borrowed values from FCTs	English	Tanzania	Ca, Fe, Zn	Food Composition	7% (27/400)	Lukmanji et al. (2008)
Zambia FCT, 2009	2009	Table	pdf	Borrowed from FCTs, analytical values from literature	English	Zambia	Ca, Fe, Zn	Item	12% (32/272)	NFNC (2009)
The Gambia FCT, 2011	2011	Table	pdf	Multiple source (analytical, borrowed, calculated)	English, but food names are mainly in Mandinka	Kiang District, Gambia	Ca, Fe, Zn	Mineral	2% (10/470)	Prynne and Paul (2011)
Mozambique FCT, 2011	2011	Table/database	pdf/xlsx	Multiple source (analytical, borrowed, calculated)	English	Mozambique	Ca, Fe, Zn	Mineral	10% (20/205)	Korkalo et al. (2011)
Central and Eastern Uganda FCT, 2012	2012	Table/database	pdf/xlsx	Borrowed values from FCTs	English	Central and Eastern Uganda	Ca, Fe, Zn	Item	4% (27/727)	Hotz et al. (2012)
Cameroon NCT, 2013	2013	Table	pdf	Analytical from literature	English	Cameroon	Ca, Fe, Se, Zn	Item	3% (4/117)	Kouebou et al. (2013)
Regional Africa NCT, 2014	2014	Database	xlsx	Borrowed values from FCT and analytical values from publication	English	Western, Eastern and Southern Africa	Ca, Fe, I, Se, Zn	Mineral	24% (66/276)	Joy et al., 2014
Mali FCT, 2015	2015	Table	pdf	Borrowed from FCTs, and analytical values from publications (Zn)	English	Mali	Ca, Fe, Zn	Mineral	40% (10/25)	Koréissi-Dembélé et al. (2017); Koreissi (2015)
Kenya FCT, 2018	2018	Table/database	pdf/xlsx	Multiple sources (analytical, borrowed, calculated)	English	Kenya	Ca, Fe, Se, Zn	Item	6% (35/663)	[dataset] FAO/Government of Kenya, 2018
Western Kenya NCT, 2019	2019	Database	xlsx	Analytical values from publications, and borrowed from FCTs	English	Western Kenya	Ca, Fe, I, Se, Zn	Mineral	22% (20/92)	Watts et al. (2019b)
Kilimanjaro Tanzania NCT, 2019	2019	Database	xlsx	Analytical values from publications, and borrowed from FCTs	English	Kilimanjaro, Tanzania	Ca, Fe, I, Se, Zn	Mineral	25% (23/92)	Watts et al., (2019a)
Malawi FCT, 2019	2019	Table	pdf	multiple sources (analytical, borrowed, calculated)	English	Malawi	Ca, Fe, I, Se, Zn	Item	8% (25/316)	van Graan et al. (2019)

(continued on next page)

Table 2 (*continued*)

Food composition table name	*Year of publication*	*Food Composition Type*	*Food Composition Format*	*Type of data*	*Language*	*Geographic Scope*	*Mineral reported*	*Reporting level*	*Food items reviewed*	*Reference*
Nigeria FCT, 2019	2019	Database	xlsx/ website	Borrowed from FCTs, and analytical values from publications	English	Nigeria	Ca, Fe, Zn	Item	10% (28/ 281)	Nigeria Food Database (2019)
Senegal NCT, 2019	2019	Table	pdf	Borrowed values from FCTs	English	Senegal	Fe, I, Zn	Mineral	25% (15/60)	Yoo et al. (2019)
Western Africa FCT, 2019	2019	Table/ database	pdf/xlsx	Multiple sources (analytical, borrowed, calculated)	English, French	Western Africa	Ca, Fe, Zn	Item	5% (56/ 1028)	Vincent et al. (2020)
Southern and Western Uganda FCT, 2021	2021	Table/ database	pdf/xlsx	Borrowed from FCTs, and product labelling	English	Southern Western Uganda	Ca, Fe, Zn	Food Composition	0	Scarpa et al. (2021)

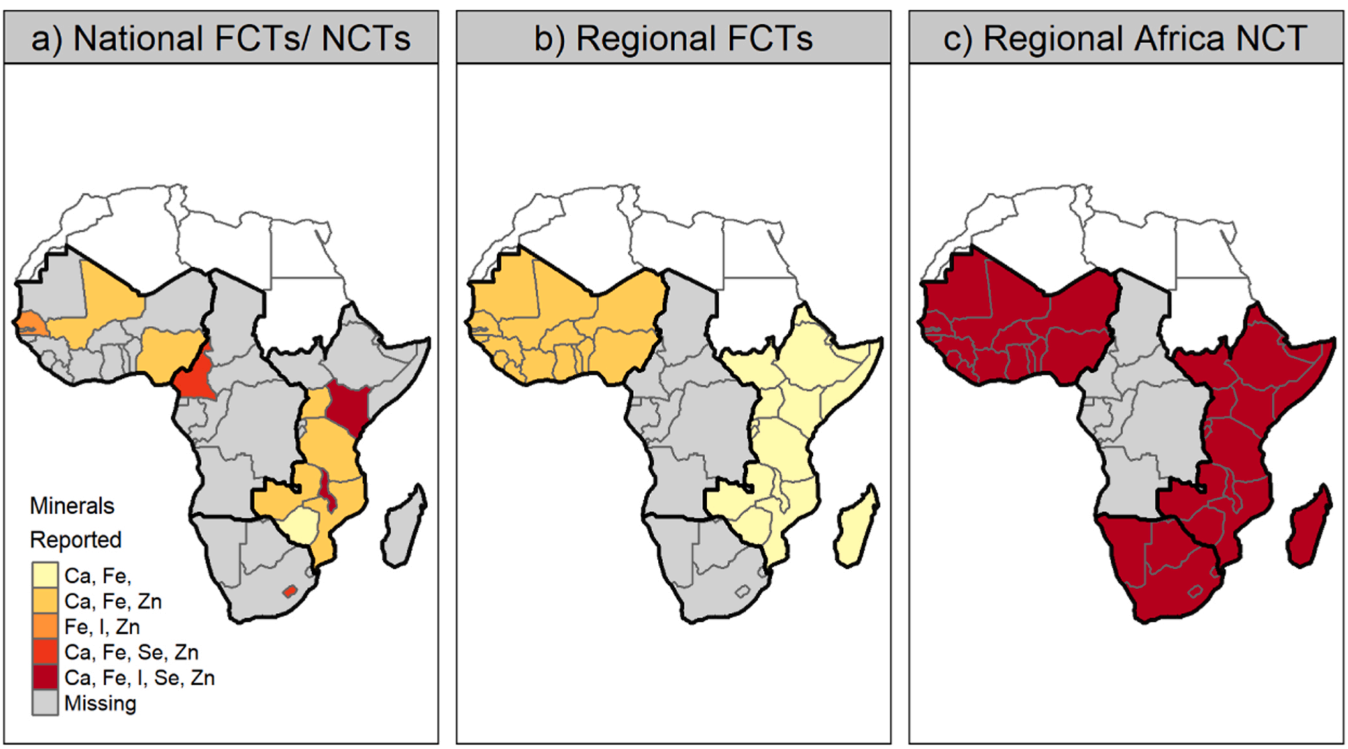


Fig. 2. Geographic scope and the minerals of interest reported in each of the FCTs/ NCTs included in the review. Panel a) shows the national food composition tables and databases/nutrient conversion tables (FCTs/NCTs) and the minerals of interest that were included. Panel b) shows the regional FCTs and the mineral of interest that were included, and panel c) shows the regional Africa NCT coverage and the mineral included.

(Hotz et al., 2012; Scarpa et al., 2021). Two regional FCTs provided information for Western Africa (Vincent et al., 2020) and Eastern Africa (West et al., 1988) (see Fig. 2.b).

The mineral nutrients that were most often included were Fe, which was reported in all the FCTs/ NCTs reviewed, Ca (reported in 18 of 19 FCTs/ NCTs) and Zn (reported in 17 FCTs/ NCTs). Conversely, I and Se were included in only five and seven FCTs/ NCTs, respectively (Table 2).

Regarding the source of information (i.e., source information for the mineral values, see Table 1), three FCTs reported it at the FCT level, eight FCTs/ NCTs reported it at food item level (of which four reported only one reference), and seven FCTs/ NCTs reported it at the mineral value level (i.e., food item nutrient level) (Table 3). Source information was only traceable, where it was reported, at the food item mineral level, or at the food item level when only one source was reported.

All the FCTs/ NCTs reviewed, except the Cameroon NCT, 2013 (Kouebou et al., 2013), reported the use of other FCTs in the compilation of the mineral data (borrowed values), whereas six FCTs and two NCTs reported the use of chemical analysis to populate the mineral values for at least a selection of foods (analytical primary values). When data were borrowed, and where it was possible to trace back to the original source, it was apparent that some FCTs/NCTs were drawing on analyses conducted many decades ago, with a heavy reliance on the Food Composition Table for use in Africa (Woot-Tsuen et al., 1968).

On average, 27 ± 17 food items per FCT/ NCT (range:0–72) were selected for assessing the percentage of missing mineral values (Section 3.3) and when mineral values were available, to identify the geographic location of the sample used to generate the mineral values (Section 3.5). The highest number of food items were extracted from the regional

Table 3
Level of reporting of the data and metadata in food composition tables and databases (FCTs) and nutrient conversion tables (NCTs) and its implications for data documentation.

Level of Reporting	Description	Implications	Food Composition Tables and Databases/ Nutrient Conversion Tables
Food composition table (FCT level)	The source of the nutrient values (sometimes referred to as bibliography or references), and other information regarding the data compiled is reported at the highest level, i.e., reporting information on the overall food composition table. For instance, authors only provided a reference list with the sources of the nutrient values for the whole food composition table.	This level of reporting precludes the identification of the source of the mineral values as well as the geographic origin of the data	Three FCTs: Zimbabwe FCT (1989), Tanzania FCT (2008), Southern and Western Uganda (2021)
Food item (food item level)	FCTs reported one or more source/s of the nutrient value per food item entry. I.e., every food item was accompanied with at least one reference stating the source of the nutrient values.	When more than one source of information was provided per food item, we could not identify which source or sources (as nutrient values are often an average of more than one value) were used when generating the mineral concentration of each food item. Hence, only for those reporting only one source of information and under the assumption that the source provided was the original source of all the reported nutrient values, could we infer the geographic origin.	Eight FCTs/NCTs reporting at this level: Four reported more than one source per food item: Cameroon NCT, 2013, Kenya FCT, 2018, Nigeria FCT, 2019, Western Africa FCT, 2019 We inferred the geographic origin and other information for four FCTs: Lesotho FCT (2006), Zambia FCT (2009), Centre and Eastern Uganda FCT (2012), Malawi FCT (2019),
Food item-nutrient (mineral level)	Information was given for every combination of food item and nutrient reported in the FCT. For instance, one (or more) references per food entry-nutrient combination were provided.	This was the only data structure that allowed us to identify the source of the nutrient values.	Seven FCTs/ NCTs: Eastern Africa FCT, 1989, Gambia FCT, 2011, Mozambique FCT, 2011, Regional Africa NCT, 2014, Mali FCT, 2015, Western Kenya NCT, 2019, Kilimanjaro Tanzania NCT, 2019, Senegal NCT, 2019

Africa NCT, 2014 (n = 72), followed by the Western Africa FCT, 2019 (n = 59). No foods were extracted from the Southern and Western Uganda FCT, 2021 and only a small number of foods were included from the Cameroon NCT, 2013 (n = 5) and The Gambia FCT, 2011 (Prynne and Paul, 2011) (n = 11), because the food items in these tables were typically cooked and/or composite dishes.

3.3. Mineral data for sub-Saharan Africa in a subset of foods

From all the FCTs/NCTs selected for review, the mineral content of 490 food items which matched the FBS food categories, were extracted. The list of food items selected and their mineral values are reported in Supplementary Table 3. In total, there were 476 values for Fe, 458 values for Ca, 383 values for Zn, 179 values for Se and 116 values for I. No values were extracted from the Mali FCT (Koréissi-Dembélé et al., 2017; Koreissi, 2015) as nutrient values were not reported.

Iodine was the mineral reported the least often (i.e., 24% of the foods). Notably, of the five FCTs/ NCTs that reported I concentration, only the regional Africa NCT and the Senegal NCT reported no missing values, whereas the other three FCTs/NCTs reported missing values, with > 60% missing values in the food entries included in this review (Fig. 3.a). None of the FCTs/NCTs reported I concentration of any pelagic fish (Fig. 3b). Of the selected foods, Se concentration values were reported for 37%, Zn was reported for 79% and Ca and Fe were reported for > 90% of the food entries reviewed.

3.4. Documentation and source of mineral concentration values

Information (metadata) on the source of each mineral value was identified for 314 (64%) food items. There were 201 food items that provided the source of information at mineral level with one unique reference per food and mineral value, except for three food items from the Mozambique FCT, 2011 (Korkalo et al., 2011) which provided two references per mineral value. The information source was provided at the food item level for 113 food items. For the other food items (n = 176), metadata were reported at the FCT level or there were multiple references provided at food item levels, precluding our ability to trace the type and geographic origin of mineral values.

The majority of the mineral values (n = 641) reported were borrowed from African FCTs as shown in Fig. 4. The second most prevalent type of data were values from FCTs outside Africa (n = 169) for all minerals except Se, for which analytical values from secondary data (e.g., peer-reviewed publications, thesis) were more frequently reported. The most common FCT source for I and Se was the Western Africa FCT, 2010 (Stadlmayr et al., 2010) whereas for Fe, Ca and Zn, it was the Tanzania FCT, 2008 (Lukmanji et al., 2008). Of the values that were borrowed from FCTs outside of Africa, the most cited sources were the Denmark FCT (National Food Institute, 2015) for I and the USDA FCT (multiple versions) for Ca, Fe, Se and Zn (U.S. Department of Agriculture, 2017; U.S. Department of Agriculture, 2005; U.S. Department of Agriculture, 2008; U.S. Department of Agriculture, 2011).

There were 110 food item-nutrient values in six FCTs/NCTs that used analytical “secondary” values, while 106 values in three FCTs/NCTs were analytical “primary” data: Kilimanjaro Tanzania NCT (2019) (Watts et al., 2019a), Western Kenya NCT (2019) (Watts et al., 2019b) and Zambia FCT (2009) (NFNC, 2009).

3.5. Geographic origin of the mineral data for sub-Saharan Africa

The geographic origin of most of the mineral values reviewed was unknown or not traceable. For those values where the geographic origin could be identified, the most cited location for Ca, Fe and Zn was the United States (US), while for I and Se it was Denmark and Malawi, respectively.

Most of the analytical secondary data used to populate FCTs/ NCTs were collected from Malawi (14 studies) and reported the sample

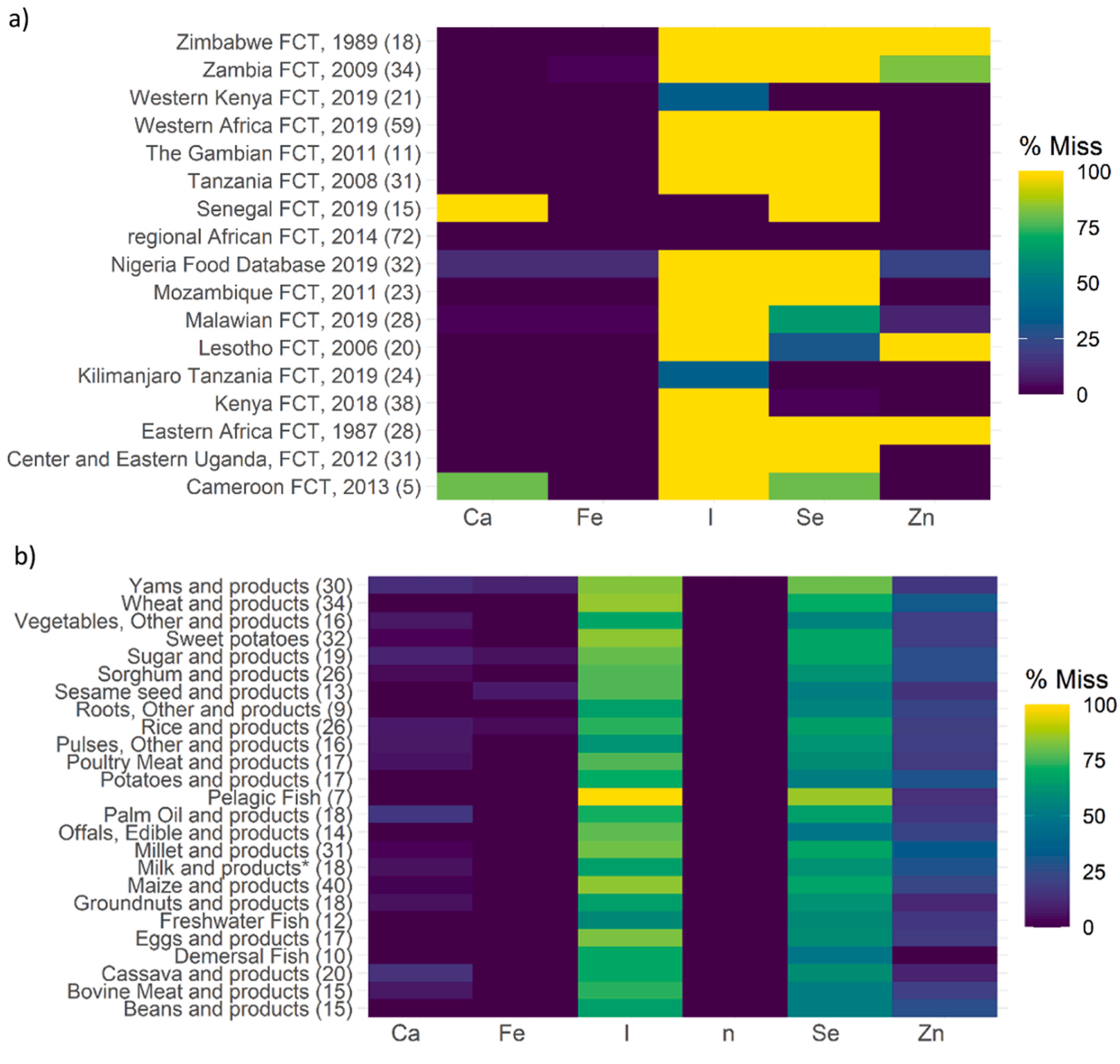


Fig. 3. Percentage of missing values per mineral of interenst in each food item. a). Percentage of missing values per mineral of interest in each FCTs reviewed. b) Percentage of missing values per mineral of interest in each food category. Number of food items per FCT/ NCT and food category are in brackets.

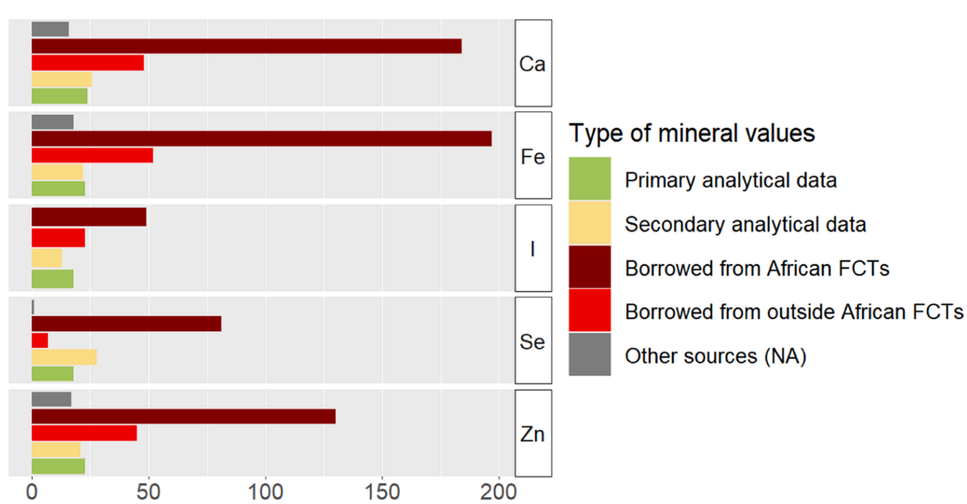


Fig. 4. Count of mineral values used to populate eleven of the Food Composition Tables and Databases/ Nutrient conversion Tables (FCTs/ NCTs) included in the review by type of value and mineral. Primary and secondary analytical values included were all from Africa.

location (only one location could not be identified because the full text was unavailable). The reporting of the sample location varied in the degree of detail, from reporting the overall location: e.g., Lilongwe, Malawi (Mumba and Jose, 2005) to providing a map with the sampling locations and associated latitude/longitude co-ordinates (Joy et al., 2015a).

Finally, the sources of only 33% (363/1084) of the mineral values were traceable to allow an identification of the geographic location of the analysed samples. As a proportion of the total number of mineral values reviewed (n = 1664), only 22% of the values from 11 of the 19 FCTs/NCTs provided sufficient information to trace the origin of food item-nutrient values.

4. Discussion

4.1. Availability of food composition tables for use in sub-Saharan Africa

Nineteen freely available FCTs/NCTs that provided Ca, Fe, I, Se and Zn food composition data, for estimating dietary intakes and supplies in sub-Saharan Africa, were reviewed. The availability of country-specific FCTs/NCTs was uneven across the region, especially for countries in the Central African region, as has been reported elsewhere (Kouebou et al., 2013). This lack of country-specific FCTs has been partially attributed to the high economic cost of compiling FCTs, the lack of funding for FCT activities, and a lack of technical and laboratory capacity for chemical analyses of micronutrients (Ene-Obong et al., 2019; Kouebou et al., 2013; Micha et al., 2018). The deficiencies of spatially-resolved food composition data and data structures have been observed previously, and the issue is not limited to sub-Saharan Africa (de Bruyn et al., 2016; Delgado et al., 2021; Ferraz de Arruda et al., 2023; Li et al., 2023; Ocké et al., 2021)). However, high-income countries generally fare better, and 75% of European countries have invested in the development of FCTs compared to only 39% of African countries, according to the food composition databases registry on the FAO/INFOODS website (INFOODS, 2020).

The importance of having country-specific FCTs have been highlighted by other authors (de Bruyn et al., 2016; Ene-Obong et al., 2019; Ocké et al., 2021). However, information on the mineral content of local foods is essential for estimating the risk of inadequate intakes of minerals to inform nutrition related policies and intervention planning to improve population health, underscoring the need for more chemical analyses of location-specific food samples in sub-Saharan African countries (Ahmad et al., 2021; Danster and Wolmarans, 2008; de Benoist et al., 2004; de Bruyn et al., 2016; Combs, J., 2015; Fuge and Johnson, 2015). Novel analyses would benefit from the improved precision and sensitivity of analytical methods and instrumentation that are currently available. Furthermore, crop mineral composition has likely changed over recent decades due to the adoption of new crop variants, and potentially due to changes in environment and management (e.g. fertilizer use) (Fan et al., 2008) and the reliance on data generated several decades ago could introduce a large amount of error to estimates of dietary mineral intakes.

4.2. Mineral data availability in selected foods

From all the food items reviewed, over 79% had values for Fe, Ca, and Zn, whereas less than 27% had values for Se and I. Lower coverage of I and Se was expected as they were included in fewer FCTs/NCTs than the other three minerals. The well-known national and sub-national geographic variability in the I and Se content of foods paired with a lack of location-specific food composition analytical values for these minerals, are common reasons for excluding I and Se from national or regional FCTs. For example, I and Se concentration values were reported in the 2010 Western African FCT but were excluded from the 2012 and 2019 editions due to these concerns (FAO et al., 2012; Stadlmayr et al., 2013; Vincent et al., 2020). Additionally, when reported, there were often missing values. For example, in the Cameroon FCT, 71% of the foods did not report Se, and in the Malawi FCT, 73% of Se and 83% of I values were missing (van Graan et al., 2019; Kouebou et al., 2013).

Since the quantities of I and Se in food items are typically 1–2 orders of magnitude lower than those of Ca, Fe, and Zn, determining their concentrations calls for more sensitive analytical approaches. This can have accessibility and cost implications, which, together with the previously mentioned geographic variability, may contribute to the high number of missing values.

A lack of relevant and accurate food composition data is problematic when attempting to characterise food system and dietary micronutrient supplies, estimate dietary micronutrient intakes, or when conducting ex-ante evaluations of nutrition interventions. For example, we were unable to identify any I concentration values for pelagic (i.e., marine) fish in the FCTs/ NCTs reviewed, although the marine environment has naturally elevated I concentrations. Fish is widely recognised as an important source of many micronutrients, particularly small fish which can be consumed whole and are often relatively affordable (Byrd et al., 2021; Hicks et al., 2019; Ryckman et al., 2021).

Equally problematic is the lack of spatially resolved mineral values, not just for Se and I, where regional differences in concentration leading to inadequate and/or excessive intakes have been extensively documented (Ahmad et al., 2021; Fuge and Johnson, 2015; Hurst et al., 2013; Ligowe et al., 2020) but also for Ca, Fe and Zn, where the spatial variability is less often characterised (Gashu et al., 2021). Thus, in some areas in sub-Saharan Africa where populations often consume locally produced foods, a single mineral value, even when sourced locally, could be inaccurate and result in misleading information for nutritional studies or programmes. However, this review found that, existing FCT data structures and compilation methods do not accommodate the integration of spatially resolved mineral composition data.

4.3. Documentation and geographic location of the mineral values

One of the aims of the scoping review was to identify the geographic location of the newly reported or borrowed mineral composition data in FCT/ NCT for use in sub-Saharan Africa by tracing it back to food samples analysed. However, the documentation sometimes did not report sufficient information to identify the source of information. In other cases, documentation was provided to identify information source which, in turn, were unavailable precluding the identification of the geographic location of the original samples used to provide mineral values. Thus, the geographic location and the relevance to the nation(s) in which the data were being used could be asserted of only 22% of the minerals in the FCTs/NCTs reviewed.

Data documentation are required to evaluate whether the nutrient values represent the nutrient content of foods in local food systems and, thus, their suitability for estimating mineral intakes of a population. Another barrier was data access in terms of where data is stored (i.e., behind paywalls, on institution hard drives, etc.) and maintained (hard copies, obsolete software formats), which again impede the use of the nutrient data for use in dietary intake estimations. Poor data documentation is one of the main barriers for evaluating data quality, reproducibility, transparency, and re-usability.

Further, some mineral values were averaged across multiple sources of food composition data, and the references were provided as a list of references either for the whole FCT, or for each food item (i.e., listing all the references for all the components together as a string). This method is recommended by some food compilation guidelines to provide representative values when sourcing nutrient concentration from single studies, and it was common practice when compiling the most recent FCTs (Western Africa FCT (2019), Nigeria FCT (2019), Kenya FCT (2018)). However, this structure (i.e., reporting average values) precludes quality assessment of individual nutrient values in the FCT, and averaged values can include data from diverse geographic locations (i.e., from India, Denmark, UK, or USA), as reported in the Western Africa or

Kenya FCTs (FAO/Government of Kenya, 2018; Vincent et al., 2020).

From the mineral values with documentation, only 33% could be traced back to the original analytical values and its geographic location. Most of the values for Ca, Fe and Zn were sourced from the US, while most of the I values were sourced from Denmark. Only for Se did the majority of values come from Africa, particularly Malawi. There is a paucity of analytical food composition data for foods grown and prepared in sub-Saharan Africa, and when available, they might not be usable because of obsolete analytical methods or insufficient documentation (Stadlmayr et al., 2013). Consequently, mineral composition data are often borrowed from other FCTs with a heavy reliance on FCTs from outside Africa (de Bruyn et al., 2016; FAO et al., 2012).

Moreover, wide variability in the mineral values of plant foods in sub-Saharan Africa has previously been reported (Barikmo et al., 2007; Stadlmayr et al., 2013). Although variation in concentrations of different minerals can be partially explained by differences in sampling and analytical methods, the variations were most likely due to crop variety and environmental factors. As highlighted by other authors, there is a need to generate local food composition data with proper data documentation (Chan et al., 2021; de Bruyn et al., 2016; Greenfield and Southgate, 2003; Lachat et al., 2016; Traka et al., 2020), and this is particularly important in countries with different agro-ecological zones, where populations consume locally sourced foods and have different traditional cooking practices, such as Nigeria (Ene-Obong et al., 2013), Cameroon (Kouebou et al., 2013), Mali (Barikmo et al., 2007, 2004), Malawi (Joy et al., 2015a), and Uganda (Scarpa et al., 2021). Hence, there is a growing recognition of the importance of spatial variability in mineral concentration in foods. For example, the US Data Central is adding two new data types (Food Foundation and Experimental Food) with a focus on nutrient variability and providing metadata that will allow users to assess the impact of that spatial variation in dietary intakes (Fukagawa et al., 2021).

4.4. Strengths and limitations

The strengths of this study included the development and publication of the study protocol and search strategy following the PRISMA-P and PRISMA-S guidelines (Moher et al., 2015; Rethlefsen et al., 2021). In addition, PRISMA-ScR guidelines (Tricco et al., 2018) were followed for reporting the results of the scoping review which were systematic including both electronic databases and other sources of information. Moreover, data and metadata were reviewed as part of the systematic approach and scripts were developed and published to increase the transparency and reproducibility of the results.

One limitation of this study was only FCTs/ NCTs that were free to access were reviewed, which may have resulted in selection bias. For example, the South Africa FCT (SAFOODS) was excluded because it is not fully open access and freely available for use and reuse, even though the SAFOODS may be of high quality (46% of values were from South African foods, and 23% were US values) and contains a high number of foods (n = 1741) (SAFOODS, 2019). However, the reason for its exclusion, i.e., it is not open access, has precluded its use previously (MoA, 2021). Similarly, 6 FCTs/NCTs were inaccessible which included FCTs that were only available as hard copies in the country's library or insufficient information (i.e., no citation or year of publication provided) prevented their identification and thus, we were unable to find them. Furthermore, those available only as hard copies that were identifiable and findable were published more than two decades ago (1957–1998). Inclusion of databases that were not freely available and accessible, would not materially change the main findings of this review in terms of data availability across the sub-Saharan Africa region, including the inadequacy of data, data structures and metadata reporting.

Another potential limitation of this study is that only raw foods were reviewed, which reduced the number of foods included from some FCTs/ NCTs. In the Southern and Western Uganda FCT all foods reported were mixed dishes, resulting in the exclusion of all data. Other FCTs/ NCTs, in which cooked foods were predominately reported, were: The Gambian FCT, the Cameroon NCT and the Centre and Eastern Uganda FCT. The inclusion of foods from the Southern and Western Uganda FCT and the Centre and Eastern Uganda FCT likely would not have changed the findings of this review, because in the former, the sources of nutrient values were only provided at the food level, and hence precluded tracing the source of the mineral values. In the latter, the primary source of data was the USDA FCT, 2008 (Hotz et al., 2012), supporting our observation that most data are either of unknown origin or not from African nations which is in line with our main results. The Gambian FCT and the Cameroon FCT reported analytical "secondary" data for some cooked food items. Additional analyses, which assessed whether the inclusion of FCTs from which few foods were reviewed would influence the results, showed the main results (i.e., most mineral values were imputed form other FCTs) did not change (Supplementary Figure 1).

Another limitation is the inclusion of specific food items which were pre-selected based on the FBS food categories. The selection of a limited number of food items was necessary to make the review feasible, and food items were chosen based on their large contributions to dietary supplies of energy and the minerals included in this review. However, FBS food categories are limited in number, which may have resulted in the omission of local foods with a high content of the minerals reviewed (Grünberger, 2014; Smith et al., 2016). Despite its limitations, FBS data have been previously used to estimate micronutrient inadequacies at global (Beal et al., 2017; Wessells et al., 2012), regional (Joy et al., 2014), and local (Watts et al., 2019) scales, because they are considered a reasonable proxy for micronutrient supply when food consumption data are not available (Coates et al., 2012).

5. Conclusion and recommendations

Food composition data are critical for dietary assessment to identify populations at risk of inadequate micronutrient intakes and inform policy and agriculture-nutrition-health intervention actions. However, this review shows the food composition data currently available for estimating dietary intakes of Ca, Fe, Zn, I, and Se, for populations in sub-Saharan Africa is limited, and rarely documents data sources at food item – nutrient level. More chemical analysis of minerals, for foods locally-grown and consumed in sub-Saharan Africa, are required, as are data structures that allow the use of spatially-relevant mineral composition data. Hence, not only should the documentation describe the analytical methods, but it should also include the geographic location of the food samples analysed. Similarly, as recommended by Greenfield and Southgate (2003), when compiling food composition from various sources, nutrient values should be annotated with sufficient information that would avoid the need of consulting the original data source. Data management and documentation could be improved by applying nutri-informatics as proposed by Chan et al., (2021). For example, the use of food and nutrition ontologies would standardise the language used in food composition allowing for the integration and interoperability of nutrition datasets (Andrés-Hernández et al., 2022). Community standards for data management should be developed to provide minimum information standards specifying essential data and documentation. All of this, together with reporting standards that would ensure comprehensive documentation would set the path to the adoption of the FAIR (Findable, Accessible, Interoperable and Reusable) principles for nutrition data (Lachat et al., 2016; Savoi et al., 2021; Top et al., 2022; Wilkinson et al., 2016).

CRediT authorship contribution statement

Lucia Segovia de la Revilla: Visualization, Project administration, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing - original draft, Writing - review & editing. **Elaine Ferguson:** Conceptualization, Methodology, Writing –

review & editing, Supervision. **Claire Dooley:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Gareth Osman:** Data Curation, Writing – review & editing. **Louise Ander:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Edward J.M. Joy:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data and code are publicly available in GitHub and OSF, linked in methods section of the paper and in the published protocol (10.31219/osf.io/vd2mf).

Acknowledgements

This work and LSdlR, EJMJ, ELF, GO and ELA were supported, in whole or in part, by the Bill & Melinda Gates Foundation (INV-002588, Micronutrient Action Policy Support (MAPS) project, additional support for CD was obtained from the Innovative Methods and Metrics for Agriculture, Nutrition and Health Actions (IMMANA) Programme, funded by UK Foreign Commonwealth and Development Office (FCDO), grant number 300654, and the Bill & Melinda Gates Foundation, grant number INV-002962. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. The funder will have no input on the design, interpretation, or publication of the study. We also would like to thank Ana Moltedo for her contribution to the definition of "Nutrient Conversion Table".

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfca.2023.105322.

References

Ahmad, S., Bailey, E.H., Arshad, M., Ahmed, S., Watts, M.J., Young, S.D., 2021. Multiple geochemical factors may cause iodine and selenium deficiency in Gilgit-Baltistan, Pakistan. Environ. Geochem Health 43, 4493–4513. https://doi.org/10.1007/s10653-021-00936-9.

Andrés-Hernández, L., Blumberg, K., Walls, R.L., Dooley, D., Mauleon, R., Lange, M., Weber, M., Chan, L., Malik, A., Møller, A., Ireland, J., Segovia, L., Zhang, X., Burton-Freeman, B., Magelli, P., Schriever, A., Forester, S.M., Liu, L., King, G.J., 2022. Establishing a common nutritional vocabulary - from food production to diet. Front. Nutr. 9.

Arsenault, J.E., Hijmans, R.J., Brown, K.H., 2015. Improving nutrition security through agriculture: an analytical framework based on national food balance sheets to estimate nutritional adequacy of food supplies. Food Secur. 7, 693–707. https://doi.org/10.1007/s12571-015-0452-y.

Barikmo, I., Ouattara, F., Oshaug, A., 2004. Protein, carbohydrate and fibre in cereals from Mali - how to fit the results in a food composition table and database. J. Food Compos. Anal. 17, 291–300. https://doi.org/10.1016/j.jfca.2004.02.008.

Barikmo, I., Ouattara, F., Oshaug, A., 2007. Differences in micronutrients content found in cereals from various parts of Mali. J. Food Compos. Anal., 6th Int. Food Data Conf. 20, 681–687. https://doi.org/10.1016/j.jfca.2007.04.002.

Beal, T., Massiot, E., Arsenault, J.E., Smith, M.R., Hijmans, R.J., 2017. Global trends in dietary micronutrient supplies and estimated prevalence of inadequate intakes. PLOS ONE 12, e0175554. https://doi.org/10.1371/journal.pone.0175554.

Belay, A., Joy, E.J.M., Chagumaira, C., Zerfu, D., Ander, E.L., Young, S.D., Bailey, E.H., Lark, R.M., Broadley, M.R., Gashu, D., 2020. Selenium deficiency is widespread and spatially dependent in ethiopia. Nutrients 12, 1565. https://doi.org/10.3390/nu12061565.

de Benoist, B., Andersson, M., Egli, I., Takkouche, B., Allen, H., 2004. Iodine status worldwide WHO Global Database on Iodine Deficiency.

Botoman, L., Chagumaira, C., Mossa, A.W., Amede, T., Ander, E.L., Bailey, E.H., Chimungu, J.G., Gameda, S., Gashu, D., Haefele, S.M., Joy, E.J.M., Kumssa, D.B., Ligowe, I.S., McGrath, S.P., Milne, A.E., Munthali, M., Towett, E., Walsh, M.G., Wilson, L., Young, S.D., Broadley, M.R., Lark, R.M., Nalivata, P.C., 2022. Soil and landscape factors influence geospatial variation in maize grain zinc concentration in Malawi. Sci. Rep. 12, 7986. https://doi.org/10.1038/s41598-022-12014-w.

de Bruyn, J., Ferguson, E., Allman-Farinelli, M., Darnton-Hill, I., Maulaga, W., Msuya, J., Alders, R., 2016. Food composition tables in resource-poor settings: exploring current limitations and opportunities, with a focus on animal-source foods in sub-Saharan Africa. Br. J. Nutr. 116, 1709–1719. https://doi.org/10.1017/S0007114516003706.

Byrd, K.A., Thilsted, S.H., Fiorella, K.J., 2021. Fish nutrient composition: a review of global data from poorly assessed inland and marine species. Public Health Nutr. 24, 476–486. https://doi.org/10.1017/S1368980020003857.

Chan, L., Vasilevsky, N., Thessen, A., McMurry, J., Haendel, M., 2021. The landscape of nutri-informatics: a review of current resources and challenges for integrative nutrition research. Database 2021. https://doi.org/10.1093/database/baab003.

Chitsiku, I.C., 1989. Nutritive Value of Foods of Zimbabwe. Zambezia vol. 16, no. 1. (pp. 67–98.) UZ, Mt. Pleasant, Harare: UZ Publications. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/6602.

Coates, J., Colaiezzi, B., Fiedler, J.L., Wirth, J., Lividini, K., Rogers, B., 2012. A program needs-driven approach to selecting dietary assessment methods for decision-making in food fortification programs. S146–S156 Food Nutr. Bull. 33. https://doi.org/10.1177/15648265120333S202.

Combs, J, F., 2015. Biomarkers of selenium status. Nutrients 7, 2209–2236. https://doi.org/10.3390/nu7042209.

Danster, N., Wolmarans, P., 2008. Characteristics of the South African Food Composition Database, an essential tool for the nutrition fraternity in the country: Part I. South Afr. J. Clin. Nutr. 21, 308–313. https://doi.org/10.1080/16070658.2008.11734172.

Delgado, A., Issaoui, M., Vieira, M.C., Saraiva de Carvalho, I., Fardet, A., 2021. Foodm composition databases: does it matter to human health? Nutrients 13, 2816. https://doi.org/10.3390/nu13082816.

Ene-Obong, H., Schönfeldt, H.C., Campaore, E., Kimani, A., Mwaisaka, R., Vincent, A., Ati, J.E., Kouebou, P., Presser, K., Finglas, P., Charrondiere, U.R., 2019. Importance and use of reliable food composition data generation by nutrition/dietetic professionals towards solving Africa's nutrition problem: constraints and the role of FAO/INFOODS/AFROFOODS and other stakeholders in future initiatives. In Proceedings of the Nutrition Society 78, 496–505. https://doi.org/10.1017/S0029665118002926.

Ene-Obong, H.N., Sanusi, R.A., Udenta, E.A., Williams, I.O., Anigo, K.M., Chibuzo, E.C., Aliyu, H.M., Ekpe, O.O., Davidson, G.I., 2013. Data collection and assessment of commonly consumed foods and recipes in six geo-political zones in Nigeria: important for the development of a National Food Composition Database and Dietary Assessment. Food Chem. 140, 539–546. https://doi.org/10.1016/j.foodchem.2013.01.102.

Fan, M.-S., Zhao, F.-J., Fairweather-Tait, S.J., Poulton, P.R., Dunham, S.J., McGrath, S.P., 2008. Evidence of decreasing mineral density in wheat grain over the last 160 years. J. Trace Elem. Med. Biol. 22, 315–324. https://doi.org/10.1016/j.jtemb.2008.07.002.

FAO. Food Balance Sheet (2010-). License: CC BY-NC-SA 3.0 IGO [dataset] . Extracted from: https://www.fao.org/faostat/en/#data/FBS. Data of Access: 25/04/2021.

FAO/Government of Kenya, 2018. Kenya food composition tables. Nairobi 254. ⟨https://www.fao.org/3/I8897EN/i8897en.pdf⟩.

FAO/Government of Kenya. 2018. Kenya Food Composition Tables. Nairobi, 254 pp [dataset] . http://www.nutritionhealth.or.ke/wp-content/uploads/Downloads/Kenya%20Food%20Composition%20Tables%20Excel%20files%202018.xlsx.

Ferraz de Arruda, H., Aleta, A., Moreno, Y., 2023. Food composition databases in the era of Big Data: vegetable oils as a case study. Front Nutr. 9, 1052934 https://doi.org/10.3389/fnut.2022.1052934.

Food and Agriculture Organization of the United Nations (FAO), 2012. International Network of Food Data Systems, West African Health Organization. In: West African food composition table / Table de composition desments d'Afrique de l'Ouest. Bioversity International, Stadlmayr, B, p. 173–p, 978–92-5–007207-4. https://hdl.handle.net/10568/104619.

Fuge, R., Johnson, C.C., 2015. Iodine and human health, the role of environmental geochemistry and diet, a review. Appl. Geochem. 63, 282–302. https://doi.org/10.1016/j.apgeochem.2015.09.013.

Fukagawa, N.K., McKillop, K., Pehrsson, P.R., Moshfegh, A., Harnly, J., Finley, J., 2021. USDA's fooddata central: what is it? and Why is it needed today? Am. J. Clin. Nutr. nqab397. https://doi.org/10.1093/ajcn/nqab397.

Gashu, D., Nalivata, P.C., Amede, T., Ander, E.L., Bailey, E.H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S.M., Hailu, K., Joy, E.J.M., Kalimbira, A.A., Kumssa, D.B., Lark, R.M., Ligowe, I.S., McGrath, S.P., Milne, A.E., Mossa, A.W., Munthali, M., Towett, E.K., Walsh, M.G., Wilson, L., Young, S.D., Broadley, M.R., 2021. The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. Nature 594, 71–76. https://doi.org/10.1038/s41586-021-03559-3.

Greenfield, H. and Southgate, D.A.T., 2003, Food composition data: Production, Management and Use, 2nd ed., FAO, Rome. Download at http://www.fao.org/infoods/publications_en.stm.

Grünberger, K., 2014. Estimating food consumption patterns by reconciling food balance sheets and household budget surveys. Rome: Food and Agriculture Organization of the United Nations.

Hicks, C.C., Cohen, P.J., Graham, N.A.J., Nash, K.L., Allison, E.H., D'Lima, C., Mills, D.J., Roscher, M., Thilsted, S.H., Thorne-Lyman, A.L., MacNeil, M.A., 2019. Harnessing global fisheries to tackle micronutrient deficiencies. Nature 574, 95–98. https://doi.org/10.1038/s41586-019-1592-6.

Hotz, C., Abdelrahman, L., Sison, C., Moursi, M., Loechl, C., 2012. A food composition table for Central and Eastern Uganda. Washington, DC: International Food Policy Research Institute and International Center for Tropical Agriculture.
Hurst, R., Siyame, E.W.P., Young, S.D., Chilimba, A.D.C., Joy, E.J.M., Black, C.R., Ander, E.L., Watts, M.J., Chilima, B., Gondwe, J., Kang'ombe, D., Stein, A.J., Fairweather-Tait, S.J., Gibson, R.S., Kalimbira, A.A., Broadley, M.R., 2013. Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi. Sci. Rep. 3, 1425–1425.
INFOODS, 2020. Tables and databases. Retrieved September 02, 2020 from: http://www.fao.org/infoods/infoods/tables-and-databases/en/.
INFOODS, 2022. Tables and databases: Africa. Retrieved November 21, 2022 from: http://www.fao.org/infoods/infoods/tables-and-databases/africa/en/.
Joy, E.J.M., Ander, E.L., Young, S.D., Black, C.R., Watts, M.J., Chilimba, A.D.C., Chilima, B., Siyame, E.W.P., Kalimbira, A.A., Hurst, R., Fairweather-Tait, S.J., Stein, A.J., Gibson, R.S., White, P.J., Broadley, M.R., 2014. Dietary mineral supplies in Africa. Table S2. Food mineral composition data from literature sources, used in conjunction with FBSs to estimate dietary mineral availability. Physiol. Plant. 151, 208–229. https://doi.org/10.1111/ppl.12144.
Joy, E.J.M., Ander, E.L., Young, S.D., Black, C.R., Watts, M.J., Chilimba, A.D.C., Chilima, B., Siyame, E.W.P., Kalimbira, A.A., Hurst, R., Fairweather-Tait, S.J., Stein, A.J., Gibson, R.S., White, P.J., Broadley, M.R., 2014. Dietary mineral supplies in Africa. Physiol. Plant. 151, 208–229. https://doi.org/10.1111/ppl.12144.
Joy, E.J.M., Broadley, M.R., Young, S.D., Black, C.R., Chilimba, A.D.C., Ander, E.L., Barlow, T.S., Watts, M.J., 2015a. Soil type influences crop mineral composition in Malawi. Sci. Total Environ. 505, 587–595. https://doi.org/10.1016/j.scitotenv.2014.10.038.
Joy, E.J.M., Kumssa, D.B., Broadley, M.R., Watts, M.J., Young, S.D., Chilimba, A.D.C., Ander, E.L., 2015b. Dietary mineral supplies in Malawi: spatial and socioeconomic assessment. BMC Nutr. 1, 42. https://doi.org/10.1186/s40795-015-0036-4.
Kihara, J., Bolo, P., Kinyua, M., Rurinda, J., Piikki, K., 2020. Micronutrient deficiencies in African soils and the human nutritional nexus: opportunities with staple crops. Environ. Geochem Health 42, 3015–3033. https://doi.org/10.1007/s10653-019-00499-w.
Koreissi, Y., 2015. Fonio (Digitaria exilis) in West Africa: towards improving nutrient quality. Fonio (Digitaria exilis) in West Africa: towards improving nutrient quality 178-pp. Thesis.
Koréissi-Dembélé, Y., Doets, E.L., Fanou-Fogny, N., Hulshof, P.J., Moretti, D., Brouwer, I. D., 2017. Comparing intake estimations based on food composition data with chemical analysis in Malian women. Public Health Nutr. 20, 1351–1361. https://doi.org/10.1017/S1368980017000015.
Korkalo, L., Hauta-alus, H., Mutanen, M., 2011. Food Compos. Tables Mozamb.: Version 2.
Kouebou, C.P., Achu, M., Nzali, S., Chelea, M., Bonglaisin, J., Kamda, A., Djiele, P., Yadang, G., Ponka, R., Ngoh Newilah, G., Nkouam, G., Teugwa, C., Kana Sop, M.M., 2013. A review of composition studies of Cameroon traditional dishes: macronutrients and minerals. Food Chem. 140, 483–494. https://doi.org/10.1016/j.foodchem.2013.01.003.
Lachat, C., Hawwash, D., Ocké, M.C., Berg, C., Forsum, E., Hörnell, A., Larsson, C., Sonestedt, E., Wirfält, E., Åkesson, A., Kolsteren, P., Byrnes, G., Keyzer, W.D., Camp, J.V., Cade, J.E., Slimani, N., Cevallos, M., Egger, M., Huybrechts, I., 2016. Strengthening the reporting of observational studies in epidemiology—nutritional epidemiology (STROBE-nut): an extension of the STROBE statement. PLoS Med. 13, e1002036 https://doi.org/10.1371/journal.pmed.1002036.
LanguaL, 2022. Links. Food Composition. Retrieved November 21, 2022 from: https://www.langual.org/langual_food_composition.asp.
Lephole, M., Khaketla, M., Monoto, M., 2006. Lesotho Food Composition Table. Maseru: Department of Agricultural Research.
Lephole, M., Khaketla, M., Monoto, M... Lesotho Food Composition Table. Maseru: Department of Agricultural Research. https://www.fao.org/fileadmin/templates/food_composition/documents/LesFCT_final_copyrevised1.xlsm.
Li, Z., Forester, S., Jennings-Dobbs, E., Heber, D., 2023. Perspective: Comprehensive Evaluation of Data Quality in Nutrient Databases. Adv. Nutr. https://doi.org/10.1016/j.advnut.2023.02.005.
Ligowe, I.S., Phiri, F.P., Ander, E.L., Bailey, E.H., Chilimba, A.D.C., Gashu, D., Joy, E.J. M., Lark, R.M., Kabambe, V., Kalimbira, A.A., Kumssa, D.B., Nalivata, P.C., Young, S. D., Broadley, M.R., 2020. Selenium deficiency risks in sub-Saharan African food systems and their geospatial linkages. Proc. Nutr. Soc. 79, 457–467. https://doi.org/10.1017/S0029665120006904.
Lukmanji, Z., Hertzmark, E., Mlingi, N.A., Assey, V., Ndossi, G., Fawzi, W., 2008. Tanzania food composition tables. MUHAS-TFNC, HSPH, Dar Es Salaam Tanzania.
Manzeke, M.G., Mtambanengwe, F., Watts, M.J., Hamilton, E.M., Lark, R.M., Broadley, M.R., Mapfumo, P., 2019. Fertilizer management and soil type influence grain zinc and iron concentration under contrasting smallholder cropping systems in Zimbabwe. Sci. Rep. 9, 6445. https://doi.org/10.1038/s41598-019-42828-0.
Micha, R., Coates, J., Leclercq, C., Charrondiere, U.R., Mozaffarian, D., 2018. Global dietary surveillance: data gaps and challenges. Food Nutr. Bull. 39, 175–205. https://doi.org/10.1177/0379572117752986.
Ministry of Agriculture (MoA). 2021. Zambia Food-Based Dietary Guidelines. Technical Report. https://www.nfnc.org.zm/download/zambia-food-based-dietary-guidelines-technical-recommendations-2021-produced-by-the-ministry-of-agriculture/.
Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., PRISMA-P Group, 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst. Rev. 4, 1. https://doi.org/10.1186/2046-4053-4-1.
Moltedo, A., Sánchez, C.Á., Troubat, N., Cafiero, C., 2018. Optimizing the use of ADePT-Food Security Module for Nutrient Analysis Version 3. Food and Agriculture Organization of the United Nations.
Mumba, P.P., Jose, M., 2005. Nutrient composition of selected fresh and processed fish species from lake Malawi: a nutritional possibility for people living with HIV/AIDS. Int. J. Consum. Stud. 29, 72–77. https://doi.org/10.1111/j.1470-6431.2005.00377.x.
National Food and Nutrition Commission (NFNC), 2009. Zambia Food Composition Tables, Fourth. Ed. Zambia. https://www.nfnc.org.zm/download/zambia-food-composition-tables-4th-edition/?wpdmdl=1500&refresh=63bee3b7b659a1673454519.
National Food Institute, 2015. Frida Fooddata (No. version 2). Technical University of Denmark, Denmark.
Neufeld, L.M., Baker, S., Garrett, G.S., Haddad, L., 2017. Coverage and utilization in food fortification programs: critical and neglected areas of evaluation. J. Nutr. 147, 1015S–1019S. https://doi.org/10.3945/jn.116.246157.
Nigeria Food Database, 2019. Retrieved November 21, 2022 from http://nigeriafooddata.ui.edu.ng/.
Nutritools, 2018. Nutritools website: useful Links. Retrieved Novemb. 21, 2022 (from). ⟨https://www.nutritools.org/links⟩.
Ocké, M.C., Westenbrink, S., van Rossum, C.T.M., Temme, E.H.M., van der Vossen-Wijmenga, W., Verkaik-Kloosterman, J., 2021. The essential role of food composition databases for public health nutrition – Experiences from the Netherlands. J. Food Compos. Anal. 101, 103967 https://doi.org/10.1016/j.jfca.2021.103967.
Page, M.J., Moher, D., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., McKenzie, J.E., 2021. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. n160 BMJ 372. https://doi.org/10.1136/bmj.n160.
Pennington, J.A., 2008. Applications of food composition data: data sources and considerations for use. J. Food Compos. Anal. 21, S3–S12. https://doi.org/10.1016/j.jfca.2007.02.005.
Pennington, J.A., Stumbo, P.J., Murphy, S.P., McNutt, S.W., Eldridge, A.L., McCabe-Sellers, B.J., Chenard, C.A., 2007. Food composition data: the foundation of dietetic practice and research. J. Am. Diet. Assoc. 107 (12), 2105–2113 https://doi.org/10.1016%2Fj.jada.2007.09.004.
Phiri, F.P., Ander, E.L., Bailey, E.H., Chilima, B., Chilimba, A.D.C., Gondwe, J., Joy, E.J. M., Kalimbira, A.A., Kumssa, D.B., Lark, R.M., Phuka, J.C., Salter, A., Suchdev, P.S., Watts, M.J., Young, S.D., Broadley, M.R., 2019. The risk of selenium deficiency in Malawi is large and varies over multiple spatial scales. Sci. Rep. 9, 6566. https://doi.org/10.1038/s41598-019-43013-z.
Popkin, B.M., Corvalan, C., Grummer-Strawn, L.M., 2020. Dynamics of the double burden of malnutrition and the changing nutrition reality. Lancet 395, 65–74. https://doi.org/10.1016/S0140-6736(19)32497-3.
Prynne, C., & Paul, A.A., 2011. Food composition table for use in The Gambia. MRC Human Nutrition Research, Cambridge.
R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/index.html.
Rethlefsen, M.L., Kirtley, S., Waffenschmidt, S., Ayala, A.P., Moher, D., Page, M.J., Koffel, J.B., Blunt, H., Brigham, T., Chang, S., Clark, J., Conway, A., Couban, R., de Kock, S., Farrah, K., Fehrmann, P., Foster, M., Fowler, S.A., Glanville, J., Harris, E., Hoffecker, L., Isojarvi, J., Kaunelis, D., Ket, H., Levay, P., Lyon, J., McGowan, J., Murad, M.H., Nicholson, J., Pannabecker, V., Paynter, R., Pinotti, R., Ross-White, A., Sampson, M., Shields, T., Stevens, A., Sutton, A., Weinfurter, E., Wright, K., Young, S., PRISMA-S Group, 2021. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst. Rev. 10 (39) https://doi.org/10.1186/s13643-020-01542-z.
Rstudio Team (2020). Rstudio: Integrated Development for R. Rstudio, PBC, Boston, MA. http://www.rstudio.com/.
Ryckman, T., Beal, T., Nordhagen, S., Chimanya, K., Matji, J., 2021. Affordability of nutritious foods for complementary feeding in Eastern and Southern Africa. Nutr. Rev. 79, 35–51. https://doi.org/10.1093/nutrit/nuaa137.
SAFOODS, 2019. South African Food Data System. Food Composition V.2019. Retrieved on November 21, 2022 from: https://safoods.mrc.ac.za/composition.html.
Savoi, S., Arapitsas, P., Duchêne, É., Nikolantonaki, M., Ontañón, I., Carlin, S., Schwander, F., Gougeon, R.D., Ferreira, A.C.S., Theodoridis, G., Töpfer, R., Vrhovsek, U., Adam-Blondon, A.-F., Pezzotti, M., Mattivi, F., 2021. Grapevine and wine metabolomics-based guidelines for FAIR data and metadata management. Metabolites 11, 757. https://doi.org/10.3390/metabo11110757.
Scarpa, G., Berrang-Ford, L., Bawajeeh, A.O., Twesigomwe, S., Kakwangire, P., Peters, R., Beer, S., Williams, G., Zavaleta-Cortijo, C., Namanya, D.B., Lwasa, S., Nowembabazi, E., Kesande, C., Rippin, H., IHACC, Team, Cade, J.E., 2021. Developing an online food composition database for an Indigenous population in south-western Uganda. Public Health Nutr. 24, 2455–2464. https://doi.org/10.1017/S1368980021001397.
Schmidhuber, J., Sur, P., Fay, K., Huntley, B., Salama, J., Lee, A., Cornaby, L., Horino, M., Murray, C., Afshin, A., 2018. The Global Nutrient Database: availability of macronutrients and micronutrients in 195 countries from 1980 to 2013. Lancet Planet. Health 2, e353–e368. https://doi.org/10.1016/S2542-5196(18)30170-0.
Segovia de la Revilla, L., Ferguson, E., Dooley, C., Osman, G., Joy, E.J.M., 2022. Assessing the availability and geographic location of food composition data used to estimate micronutrient intake in sub-Saharan. Afr.: Protoc. a scoping Rev. OSF Prepr. Dec. 9. https://doi.org/10.31219/osf.io/vd2mf.

Smith, M.R., Micha, R., Golden, C.D., Mozaffarian, D., Myers, S.S., 2016. Global expanded nutrient supply (GENuS) model: a new method for estimating the global dietary supply of nutrients. PLoS One 11, e0146976. https://doi.org/10.1371/journal.pone.0146976.

Stadlmayr, B., Charrondiere, U.R., Burlingame, B., 2013. Development of a regional food composition table for West Africa. Food Chem. 140, 443–446. https://doi.org/10.1016/j.foodchem.2012.09.107.

Stadlmayr, B., Charrondiere, U.R., Addy, P., Samb, B., Enujiugha, V.N., Bayili, R.G., Fagbohoun, E.G., Smith, I.F., Thiam, I., Burlingame, B., 2010. Composition of selected foods from West Africa. Food and Agriculture Organization, Rome, 13–14. https://www.fao.org/3/i1755e/i1755e.pdf.

Top, J., Janssen, S., Boogaard, H., Knapen, R., Şimşek-Şenel, G., 2022. Cultivating FAIR principles for agri-food data. Comput. Electron. Agric. 196, 106909 https://doi.org/10.1016/j.compag.2022.106909.

Traka, M.H., Plumb, J., Berry, R., Pinchen, H., Finglas, P.M., 2020. Maintaining and updating food composition datasets for multiple users and novel technologies: current challenges from a UK perspective. Nutr. Bull. 45, 230–240. https://doi.org/10.1111/nbu.12433.

Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., Moher, D., Peters, M.D.J., Horsley, T., Weeks, L., Hempel, S., Akl, E.A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M.G., Garritty, C., Lewin, S., Godfrey, C.M., Macdonald, M.T., Langlois, E.V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tunçalp, Ö., Straus, S.E., 2018. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann. Intern Med 169, 467–473. https://doi.org/10.7326/M18-0850.

U.S. Department of Agriculture, 2017. USDA Food Composition Databases. National Agricultural Library. Version 3.7.1. https://fdc.nal.usda.gov/.

U.S. Department of Agriculture, Agricultural Research Service, 2005. USDA National Nutrient Database for Standard Reference, Release 18. Methods and Application of Food Composition Laboratory Home Page, http://www.ars.usda.gov/nea/bhnrc/mafcl.

U.S. Department of Agriculture, Agricultural Research Service, 2008. USDA National Nutrient Database for Standard Reference, Release 21. Beltsville, MD: USDA. www.ars.usda.gov/Services/docs.htm?docid=18880.

U.S. Department of Agriculture, Agricultural Research Service, 2011. USDA National Nutrient Database for Standard Reference, Releases 21–24. http://www.nal.usda.gov/fnic/foodcomp/search/.

van Graan, A., Chetty, J., Jumat, M., Masangwi, S., Mwangwela, A., Phiri, F.P., Ausman, L.M., Ghosh, S., Marino-Costello, E. (Eds.), 2019. MAFOODS. 2019. Malawian Food Composition Tables, 1st ed. Lilongwe, Malawi. ISBN: 978–1-928340–41-6. https://dl.tufts.edu/downloads/g158bw806?filename=d217r336d.pdf.

Vincent, A., Grande, F., Compaoré, E., Amponsah Annor, G., Addy, P.A., Aburime, L.C., Ahmed, D., Bih Loh, A.M., Dahdouh Cabia, S., N, D., Dembélé, F.M., Dieudonné, B., Edwige, O.B., Ene-Obong, H.N., Fanou Fogny, N., Ferreira, M., Omaghomi Jemide, J., Kouebou, P.C., Compaore, M., Nájera Espinosa, S., Ouattara, F., Rittenschober, D., Schönfeldt, H., Stadlmayr, B., van Deventer, M., Razikou Yiagnigni, A. & Charrondière, U.R., 2020. FAO/INFOODS Food Composition Table for Western Africa (2019) – Table de composition des aliments FAO/INFOODS pour l'Afrique de l'Ouest (2019) [dataset]. https://www.fao.org/fileadmin/user_upload/faoweb/2020/WAFCT_2019.xlsx.

Vincent, A., Grande, F., Compaoré, E., Amponsah Annor, G., Addy, P.A., Aburime, L.C., Ahmed, D., Bih Loh, A.M., Dahdouh Cabia, S., N, D., Dembélé, F.M., Dieudonné, B., Edwige, O.B., Ene-Obong, H.N., Fanou Fogny, N., Ferreira, M., Omaghomi Jemide, J., Kouebou, P.C., Compaore, M., Nájera Espinosa, S., Ouattara, F., Rittenschober, D., Schönfeldt, H., Stadlmayr, B., van Deventer, M., Razikou Yiagnigni, A. & Charrondière, U.R., 2020. FAO/INFOODS Food Composition Table for Western Africa (2019) User Guide & Condensed Food Composition Table/ Table de composition des aliments FAO/INFOODS pour l'Afrique del'Ouest (2019) Guide d'utilisation & table de composition des aliments condensée. Rome, FAO. https://www.fao.org/3/ca7779b/CA7779B.PDF.

Watts, M.J., Middleton, D.R.S., Marriott, A.L., Humphrey, O.S., Hamilton, E.M., Gardner, A., Smith, M., McCormack, V.A., Menya, D., Munishi, M.O., Mmbaga, B.T., Osano, O., 2019. Source apportionment of micronutrients in the diets of Kilimanjaro, Tanzania and Counties of Western Kenya, 14447–14447. Sci. Rep. 9 https://doi.org/10.1038/s41598-019-51075-2.

Watts, M.J., Middleton, D.R.S., Marriott, A.L., Humphrey, O.S., Hamilton, E.M., Gardner, A., Smith, M., McCormack, V.A., Menya, D., Munishi, M.O., Mmbaga, B.T., Osano, O., 2019a. Source apportionment of micronutrients in the diets of Kilimanjaro,Tanzania and Counties of Western Kenya. Supplementary Table 8: Summary crop composition data by Food Group for Tanzania, 14447–14447 Sci. Rep. 9. https://doi.org/10.1038/s41598-019-51075-2.

Watts, M.J., Middleton, D.R.S., Marriott, A.L., Humphrey, O.S., Hamilton, E.M., Gardner, A., Smith, M., McCormack, V.A., Menya, D., Munishi, M.O., Mmbaga, B.T., Osano, O., 2019b. Source apportionment of micronutrients in the diets of Kilimanjaro,Tanzania and Counties of Western Kenya. Supplementary Table 9: Summary crop composition data by Food Group for Kenya, 14447–14447 Sci. Rep. 9. https://doi.org/10.1038/s41598-019-51075-2.

Wessells, K.R., Singh, G.M., Brown, K.H., 2012. Estimating the global prevalence of inadequate zinc intake from national food balance sheets: effects of methodological assumptions. PLoS One 7, e50565. https://doi.org/10.1371/journal.pone.0050565.

West, C.E., Pepping, F., Temalilwa, C.R., 1988. The composition of foods commonly eaten in East Africa. Wageningen Agricultural University.

White, J.M., Beal, T., Arsenault, J.E., Okronipa, H., Hinnouho, G.-M., Chimanya, K., Matji, J., Garg, A., 2021. Micronutrient gaps during the complementary feeding period in 6 countries in Eastern and Southern Africa: a Comprehensive Nutrient Gap Assessment. Nutr. Rev. 79, 16–25. https://doi.org/10.1093/nutrit/nuaa142.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. https://doi.org/10.1038/sdata.2016.18.

Wood, S.A., Tirfessa, D., Baudron, F., 2018. Soil organic matter underlies crop nutritional quality and productivity in smallholder agriculture. Agric., Ecosyst. Environ. 266, 100–108. https://doi.org/10.1016/j.agee.2018.07.025.

Woot-Tsuen, W.L., Busson, F., Jardin, C., 1968. Food composition table for use in Africa. United States Department of Health. Educ. Welf. Nutr. Div. 36.

World Nutrient Databases for Dietary Studies (WNDDS),2022. Retrieved November 21, 2022 from: https://foodsystems.org/resources/wndds/.

Yoo, Y.M., Atkin, R.A., Pachón, H., 2019. Short Communication: Development of a food composition table to analyze Senegalese food expenditure data. African Journal of Food, Agriculture, Nutrition and Development 19. https://doi.org/10.4314/ajfand.v19i2.

4 Chapter 4: An open science framework and tools to compile food composition data for nutrition

This chapter presents a framework and the tools to compile reproducible, reusable, efficient, and transparent Food Composition Tables and Databases (FCTs) and Nutrient Conversion Tables (NCTs) for nutrition.

The study was a collaboration with the Food and Nutrition Division of the UN Food and Agriculture Organization. It was published in the *Journal of Food Composition of Analysis* and can be accessed here (https://doi.org/10.1016/j.jfca.2024.106894).

The repository to replicate the analysis is publicly available here (https://doi.org/10.5281/zenodo.14265103). In addition, an R package was developed in collaboration with the co-authors that compiled most of the tools developed for the framework: *NutritionTools (https://doi.org/10.5281/zenodo.14193766)*.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	2006000	**Title**	Ms
First Name(s)	Lucia		
Surname/Family Name	Segovia de la Revilla		
Thesis Title	Geospatial analysis of food composition data to estimate population micronutrient intake		
Primary Supervisor	Edward Joy		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Journal of Food Composition and Analysis		
When was the work published?	November 2024		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conceptualised, performed the analysis and was the primary writer. Thomas Codd and Liberty Mlambo helped writing the code and some functions. Fernanda Grande, Doris Rittenschober, Ana Moltedo and Bridget A. Holmes helped reviewing the methodology and reviewing manuscript. E. Louise Ander and Edward J.M. Joy helped with the conceptialisation and reviewing the manuscript.

SECTION E

Student Signature	Lucia Segovia de la Revilla
Date	29/10/2024

Supervisor Signature	EJMJ - Edward Joy
Date	06/01/2025

Evidence of retention of copyright

"An open science framework and tools to create reproducible food composition data for use in nutrition"

In line with the Elsevier OA with CC-BY, I have the right to re-use the article without permission from Elsevier, see the Copyright page for more information.

Authors' rights in the article

	OA with CC BY	OA with CC BY-NC	OA with CC BY-NC-ND	Subscription
Receive proper attribution and credit for their published work	Yes	Yes	Yes	Yes
Re-use their article in their own new works, without permission from Elsevier or payment to Elsevier, including by: • making copies of the article (or part of the article) to promote companies or products they own, whether or not such promotion is commercial; • including the article in a thesis or dissertation; • extending the article to a book, including the article in a subsequent compilation of their own work, or re-using portions, excerpts, and their own figures, tables and images from the article in their own new works (which in each case may be published with Elsevier or with a third party commercial or non-commercial publisher, at the author's discretion)	Yes	Yes	Yes	Yes[1]

Journal of Food Composition and Analysis 137 (2025) 106894

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

An open science framework and tools to create reproducible food composition data for use in nutrition

Lucia Segovia de la Revilla [a,b,*], Thomas Codd [b], Edward J.M. Joy [a,e], Liberty Mlambo [b], Fernanda Grande [c], Doris Rittenschober [c], Ana Moltedo [d], Bridget A. Holmes [c], E. Louise Ander [b,f]

[a] Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom
[b] School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, United Kingdom
[c] Food and Nutrition Division, Food and Agriculture Organization of the United Nations (FAO), Viale delle Terme di Caracalla, Rome 00153, Italy
[d] Statistics Division, Food and Agriculture Organization of the United Nations (FAO), Viale delle Terme di Caracalla, Rome 00153, Italy
[e] Sustainable Soils and Crops, Rothamsted Research, West Common, Harpenden, AL5 2JQ, United Kingdom
[f] Centre for Environmental Geochemistry, British Geological Survey, Keyworth, Nottingham NG12 5GG, United Kingdom

ARTICLE INFO

Keywords:
Food composition data
INFOODS
Findability
Accessibility
Interoperability
Reproducibility
Fish and aquatic products
Metadata
Micronutrients

ABSTRACT

Food composition tables and databases (FCTs) and Nutrient Conversion Tables (NCTs) are essential for nutrition research. Compiling a new NCT requires multiple FCTs, usually with incompatible formats. FCT cleaning and standardisation is rarely reproducible and requires significant resources. Our aim was to develop a framework and tools for compilation and reporting of reproducible FCTs/NCTs, through expanding the fish and other aquatic products in the global NCT for the Food and Agriculture Organization of the United Nations (FAO) Supply and Utilization Accounts.

FAO/ International Network of Food Data Systems (INFOODS) guidelines, and open science tools were used for processing. New R functions and scripts were developed to: import and standardise 12 FCTs; re-calculate food components; perform quality checks; and format outputs (e.g., spreadsheets).

This resulted in the expansion of the global NCT, providing information on 32 food components for 95 fish and other aquatic products. The workflow takes 160 s to run. The scripts are publicly available in GitHub, with a manual, and can be used or adapted.

These open science tools provide a novel resource to create, update and expand FCTs/NCTs in a reproducible, reusable, efficient, and transparent manner, for use in nutrition research. food composition data for nutrition research.

1. Introduction

Food Composition Tables and Databases (FCTs) contain data on the energy and nutrient composition of food items, and sometimes other information such as recipes or edible portions. FCTs have multiple uses in public health nutrition, including their integration with information on food consumption to estimate intake of energy and nutrients, for example to conduct population nutrition surveillance or to explore the relationship between intake of nutrients and health outcomes (Durazzo and Lucarini, 2022; Traka et al., 2020). Most FCTs are compiled by national authorities or research groups, and they may be tailored to specific contexts or studies. Thus, these datasets are found in various formats, with differences in the list of food items and components (i.e. nutrients) reported, inconsistent use of data conventions (e.g., nutrient definition, analytical methods, mode of expression, units), and with variable quality and completeness of data, metadata and documentation (Clancy et al., 2015; Pennington et al., 2007; Segovia de la Revilla et al., 2023).

National/regional FCTs are often combined to generate study or research specific food composition datasets, for example those used for Household Consumption and Expenditure Surveys: referred to as Nutrient Conversion Tables (NCTs) from this point onwards. Compiling

* Correspondence to: London School of Hygiene and Tropical Medicine, Department of Population Health, Faculty of Epidemiology & Population Health, Keppel St, London WC1E 7HT, United Kingdom.

E-mail addresses: Lucia.Segovia-De-La-Revilla@lshtm.ac.uk (L. Segovia de la Revilla), Thomas.Codd@nottingham.ac.uk (T. Codd), Edward.joy@lshtm.ac.uk (E.J.M. Joy), Liberty.Mlambo@nottingham.ac.uk (L. Mlambo), Fernanda.Grande@fao.org (F. Grande), Doris.Rittenschober@fao.org (D. Rittenschober), Ana.Moltedo@fao.org (A. Moltedo), Bridget.Holmes@fao.org (B.A. Holmes), Louise.Ander@nottingham.ac.uk (E.L. Ander).

https://doi.org/10.1016/j.jfca.2024.106894
Received 5 August 2024; Received in revised form 11 October 2024; Accepted 23 October 2024
Available online 28 October 2024

locally relevant FCTs/NCTs requires multiple data inputs, typically including non-standardised and non-interoperable FCTs resulting in manual data processing, which is rarely reproducible, prone to human error and involves significant human and financial resources (Charrondiere et al., 2023; Durazzo et al., 2022; Zeb et al., 2021). The often inscrutable decisions made (e.g., converting trace to zero, imputing values, etc.) when generating both FCTs and NCTs may decrease the reliability and accuracy of the food composition values reported. Furthermore, many countries do not have a national FCT and/or local analytical food composition data. In such cases data are borrowed from other countries, reducing further the relevance and accuracy of the data (Bruyn et al., 2016; Ene-Obong et al., 2019; Segovia de la Revilla et al., 2023). Consequently, when used along with other datasets (e.g., food consumption) may inadvertently propagate these inconsistencies/inaccuracies into the estimates of nutrient intakes and risk of (in)adequacy (Coates et al., 2017; Joy and Kumssa, 2022; Kapsokefalou et al., 2019).

Increased data comparability and interchangeability can result from the use of standardised methods, including data collation, analysis, food description and formats, to compile FCTs (Durazzo et al., 2022; Ene-Obong et al., 2019; Ispirova et al., 2017; Kapsokefalou et al., 2019). For instance, the use of International Network of Food Data Systems (INFOODS) food component identifiers (also known as tagnames) provides unambiguous component identification facilitating data interchange (FAO/INFOODS, 2012a; Klensin et al., 1989). More recently, the development of the Compositional Dietary Nutrition Ontology (CDNO) serves a similar purpose (Andrés-Hernández et al., 2022). Likewise, data harmonisation is essential for reconciling diverse data sources and to allow for compatibility and comparability among them (Zeb et al., 2021). For example, a standardised and comprehensive food description is needed for accurate food matching when linking two different lists of foods by their descriptions (Moshfegh et al., 2022). A number of standards are currently available for use as food classification and description systems, for instance FoodEx2 classification and description system (European Food Safety Authority, 2015), LanguaL (Møller and Ireland, 2018), or using food ontologies, such as FoodOn (Dooley et al., 2018; Ispirova et al., 2017).

Sector specific classification systems are also available, such as the Aquatic Sciences and Fisheries Information System (ASFIS) (FAO, 2022) which is curated and maintained by FAO and provides consistent classification systems for fishery and aquaculture products, including grouping and identification codes (e.g., International Standard Statistical Classification for Aquatic Animals and Plants (ISSCAAP), taxonomic and 3-alpha group) and taxonomic information (e.g., scientific name, species family, etc.). Harmonised food description and classification facilitate the incorporation of different data sources, including data from different countries, while enabling the aggregation of similar foods, such as the 95 fish and other aquatic products in the FAO Supply Utilisation Accounts (SUAs) (referred in this study as “SUA items”) (FAO, 2023, 2021; Grande et al., 2024; Rittenschober et al., 2016). Hence, the use of community standards that align with the Findability, Accessibility, Interoperability and Reusability (FAIR) principles would contribute to increased transparency in the nutrition field (Chan et al., 2021).

The aim of this study was to develop an open science framework and tools to compile transparent and reproducible FCTs and NCTs. The adoption of open science approaches including publishing the code would aid other researchers and food/nutrition composition compilers to apply this framework, similar to recent efforts to harmonise food consumption data processing (Luo et al., 2021). The objectives were to combine FCTs from multiple formats by providing standardising and harmonising scripts; reduce costs of updating and generating new FCTs/NCTs; increase reproducibility, re-usability, efficiency and transparency of FCTs/NCTs. Finally, these were applied to: a) the validation of the framework by replicating the compilation of the fish and aquatic products subset of the global NCT for FAO SUAs developed by FAO’s Food and Nutrition Division (Grande et al., 2024); b) the extension of the nutrients included for the fish and other aquatic products to showcase its implementation (Fig. 1).

2. Methods

The main framework steps were: identifying and obtaining the food composition data (i.e., FCTs), standardising the food composition data into a common data library, harmonisation of the data (including food matching), checking quality and completeness of the food composition data, and compiling the FCT/NCT and relevant documentation (Fig. 2). The framework was based on the recommendations outlined in the FAO/INFOODS Guidelines (FAO/INFOODS, 2012b, 2012a) and in the Micronutrient Action Policy Support (MAPS) project scripting approaches. The steps outlined here were developed in RStudio version 2023.6.0.421 powered by the R software version 4.4.1 (Posit team, 2023; R Core Team, 2023).

2.1. Identifying and obtaining the food composition tables and databases

The selection criteria for FCTs have been fully documented elsewhere (Grande et al., 2024). In brief, FCTs of high quality were selected based on scoring undertaken using the “FAO/INFOODS Evaluation framework to assess the quality of published food composition tables and databases” (Charrondiere et al., 2023). Then, those FCTs that passed the screening were checked for relevancy for the study/context, data availability and missing values (e.g., relevant foods and nutrients are reported), and data quality and reporting (e.g., method of chemical analysis, complete metadata). After the FCTs were reviewed and selected, access to the data was obtained, when possible, in a text or tabular format (.csv, MS Excel, MS Access,.txt), and imported into R/RStudio (Table 1).

2.2. Importing the data

FCTs were found in a variety of data structures/formats which influenced the steps and complexity of importing the original FCT files into R(Studio) (see Table 1). Thus, individual R scripts were developed to perform the importing and subsequent FCT-specific steps. In addition, within our framework, a template is provided which includes guidance on several steps (e.g., choosing the import function according to the data format), and operations that are commonly required for cleaning and standardising FCTs. This is designed to facilitate script re-use for future incorporation of new datasets.

2.3. Data cleaning and standardisation

After importing the FCTs, the most frequent cleaning and standardisation steps, are described here. These are needed for compilation of the different FCTs into one food composition data library.

2.3.1. Formatting FCTs into a tabular format

The first step is getting the FCT into a tabular format, which helps with further processing, as functions can be applied across multiple foods and/or food components. Some of the formatting tasks include: removing empty rows, translocating and relocating columns, and/or merging multiple data tables (e.g., when nutrients were separated in different spreadsheets).

2.3.2. Renaming variables

Renaming variables is important for compilation as variables reporting values of the same food component should have exactly the same name. Our framework uses the FAO/INFOODS food component identifiers (tagnames), denoted by “< >” (e.g. <ENERC>) in this document (FAO/INFOODS, 2012b; Klensin et al., 1989) to precisely identify all of the food components, while we propose other common names for the remaining variables (e.g., food identifier (*fcd_id*), food

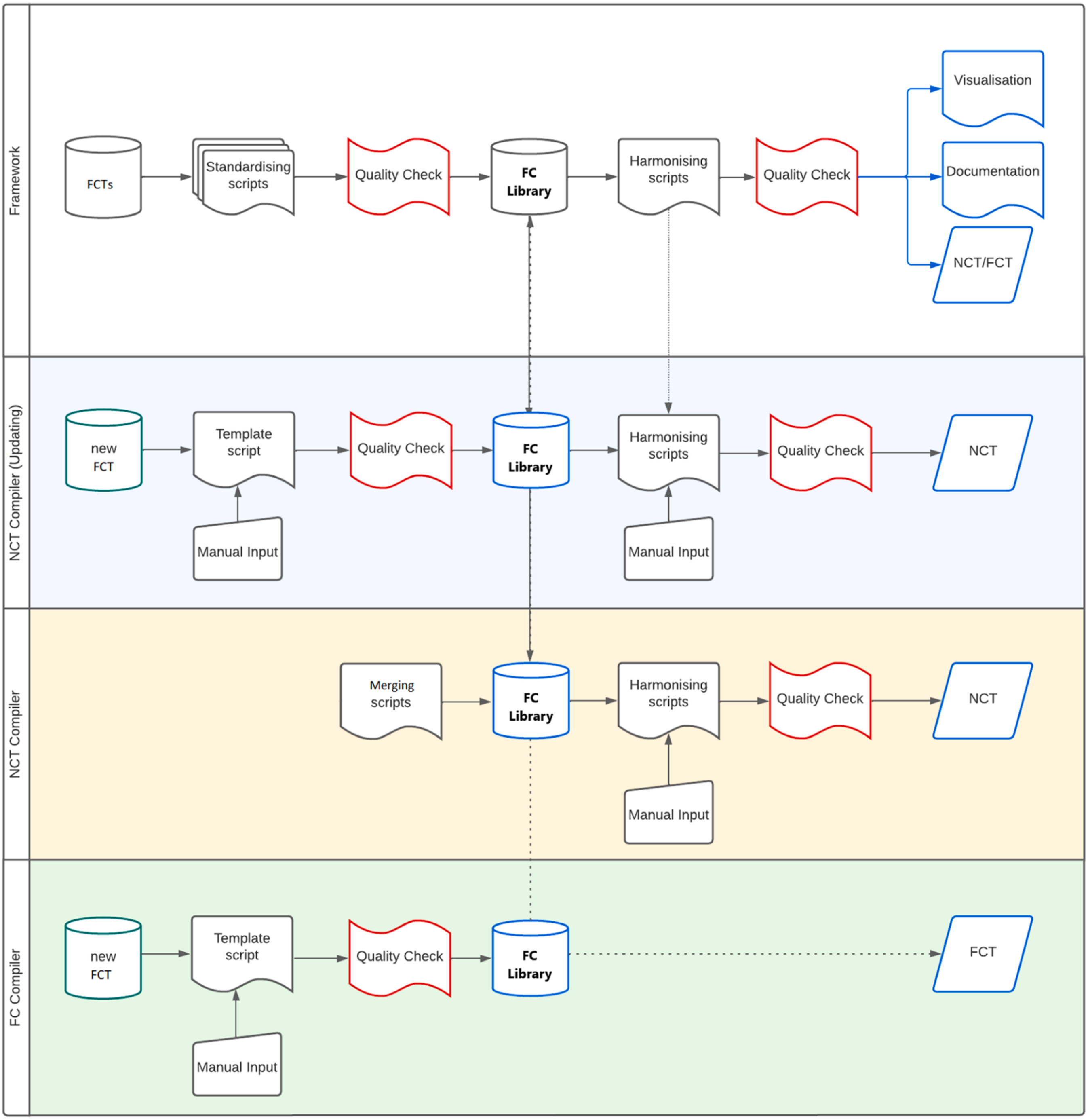


Fig. 1. The Framework (white background) presents the complete data workflow from the extraction of the "raw" Food Composition Tables (FCTs) to the standardisation into a Food Composition (FC) Library and harmonisation to the final generation of Nutrient Conversion Table (NCT), documentation and data visualisation. The NCT Compiler (updating) (blue background) presents the data journey of a user who aims to update any NCT. The NCT Compiler (yellow background) presents the data journey of a user who aims to use the standardised FCTs generated within this framework to produce an NCT. The FC Compiler (green background) presents the journey of user that aims to standardise a new FCT that could be integrated into the overall framework. The arrows across the lanes represent the steps of the workflow where users could benefit from or contribute to the workflow. The Food Composition (FC) data Library refers to one or more standardised Food Composition Tables (FCTs) which are or can be part of the Food Composition (FC) data Library generated in this study.

description (*food_desc*), etc.) (Supplementary Table 1A).

2.3.3. Standardisation of values

Firstly, special characters (e.g., "*", "[]", "-") and character strings (e.g., "trace", "LOD", "N") which are often used within FCTs need to be converted into numeric values to allow mathematical operations. For instance, values displayed alongside special characters (such as, "[45.6]") usually denote "low quality values" and/or a different method of analyses/ tagname (e.g., "Total fat using a mixed solvent extraction vs Soxhlet method"). The special characters are removed and, if a suitable tagname is available, the values are reported in a new column with the appropriate tagname. Furthermore, documentation is generated and added as metadata indicating, for instance, "low quality value" for a given food entry and component. This is stored in a new variable (called "*comments*") to retain full transparency in the data processing decision steps and for informing end-users. Similarly, for special characters and/

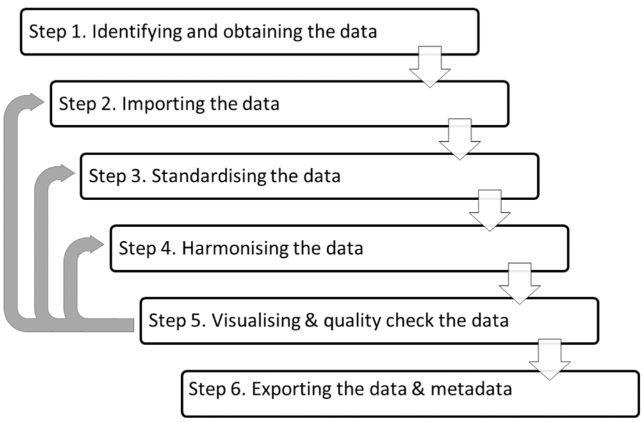


Fig. 2. The six main steps framework: 1) identifying and obtaining the data (e.g., "raw" Food Composition Tables (FCTs)), 2) importing the "raw" FCTs into R, 3) cleaning and standardisation of the FCTs into a common food composition library, 4) harmonisation of the data (including food matching), 5) checking quality and completeness of the food composition data, and 6) compiling the Nutrient Conversion Table (NCT) and relevant documentation and exporting in a standard format (e.g. Microsoft Excel, Word, etc.). Grey arrows indicate iterative feedback.

or strings used to indicate missing values and trace or below limits of detection, information is added to the "*comments*" variable, and characters are transformed to "*NA*" and zero, respectively.

2.3.4. *Units of measurements*

Finally, food components are occasionally expressed using different units (e.g., g of calcium per 100 g of fresh weight, edible portion, or mg of calcium per 100 g of fresh weight, edible portion) or denominators (e.g. per 100 g of fresh weight or per 100 g of fatty acids) between FCTs. Both units and denominators were standardised following the "FAO/INFOODS Guidelines for Converting Units, Denominators and Expressions" (FAO/INFOODS, 2012c).

2.4. *Data compilation and harmonisation*

After the FCTs are standardised, they can be compiled into a single food composition data library because they share the same variable names, units, and structure. The next sections outline the steps proposed for harmonisation and evaluation of FCT/NCTs.

2.4.1. *Food description classification/harmonisation and food matching*

One critical and time-consuming step in generating NCTs is food matching, which is the process of linking a food item (or group of food items) with the relevant foods described in the FCT. In our case study,

Table 1
List of the Food Composition Tables and Databases (FCTs) standardised and compile in the food composition data library used to validate and expand the energy and nutrient information for 95 fish and other aquatic products, including number of food items, the coverage, and data provenance.

FCT id.	FCT name	Reference	Food items (N)	Coverage	Access	Format (version)	Link to the original data
US19	USDA National Nutrient Database for Standard Reference, Legacy Release	United States Department of Agriculture USDA, 2019 (1)	7793	The United States of America	Publicly available	MS Access (legacy release)	https://agdatacommons.nal.usda.gov/articles/dataset/USDA_National_Nutrient_Database_for_Standard_Reference_Legacy_Release/24661818
AU19	Australian Food Composition Database	FSANZ Food Standards Australia New Zealand, 2019(2)	1534	Australia	Restricted-use licence[(1)]	MS Excel (version 1)	https://www.foodstandards.gov.au/science-data/monitoringnutrients/afcd/australian-food-composition-database-download-excel-files
NZ18	New Zealand Food Composition Database	New Zealand Institute for Plant and Food Research Limited & Ministry of Health, 2019 (3)	2767	New Zealand	Publicly available	MS Excel (version 01)	http://www.foodcomposition.co.nz/foodfiles
DK19	Frida: Food Database	DTU Technical University of Denmark Food Institute, 2019 (4)	1186	Denmark	Publicly available	MS Excel (version 4)	https://frida.fooddata.dk
WA19	FAO/INFOODS Food Composition Table for Western Africa	(Vincent et al., 2020) (5)	1028	Western Africa region	Publicly available	MS Excel (-)	https://www.fao.org/fileadmin/user_upload/faoweb/2020/WAFCT_2019.xlsx
KE18	Kenya Food Composition Tables	FAO & Government of Kenya, 2018 (6)	658	Kenya	Publicly available	MS Excel (-)	https://nutritionhealth.or.ke/programmes/healthy-diets-physical/food-composition-tables/
IN17	Indian Food Composition Tables	Longvah et al., 2017 (7)	528	India	Restricted-use licence[(2)]	MS Excel (-)	Not available
JA15	Standard Tables of Food Composition in Japan	MEXT, 2015 (8)	2191	Japan	Publicly available	MS Excel (version 7)	http://www.mext.go.jp/a_menu/syokuhinseibun/1365451.htm/
BA13	Food Composition Table for Bangladesh	Shaheen et al., 2013 (9)	381	Bangladesh	Publicly available	MS Excel (-)	https://www.fao.org/fileadmin/templates/food_composition/documents/FCDB_7_4_14.xlsx
BR11	Brazilian FCT (TACO)	NEPA-UNICAMP Núcleo de Estudos e Pesquisas em Alimentação – Universidade Estadual de Campinas, 2011(10)	597	Brazil	Restricted-use licence	MS Excel (version 4)	http://www.nepa.unicamp.br/arquivo/uploads/taco-4a-edicao/taco-4a-edicao-2/
UF16	FAO/INFOODS Global Food Composition Database for Fish and Shellfish (uFiSh)	FAO, 2016 (11)	515	Global	Publicly available	MS Excel (version 1.0)	https://www.fao.org/fileadmin/templates/food_composition/documents/uFiSh1.0.xlsx /
NO21	The Norwegian Food Composition Table	Norwegian Food Safety Authority, 2021(12)	2070	Norway	Publicly available	MS Excel (-)	https://www.matportalen.no/verktoy/the_norwegian_food_composition_table/
Footnote	[(1)]The file is restricted because the version 1 has been replaced with the new release (v.2). [(2)]Only publicly available in pdf which is currently not available.						

the original fish and other aquatic products in all selected FCTs, except for one, the Norwegian FCT (2021), were previously matched with their corresponding SUA item by experts from FAO's Food and Nutrition Division, as part of the compilation of the Global NCT for SUA. A detailed description of the methodology and principles applied for the food matching are documented elsewhere (Grande et al., 2024). In brief, the highest quality (i.e., the highest similarity between the SUA item reported and the FCT item description) in raw form of the food was matched, unless specified as "prepared" in the SUA item description. For fisheries and aquatic products, the scientific names were used to classify and identify the food items using the ISSCAAP code and 3-alpha codes which together with the food name description aided the food matching process.

The Norwegian FCT (2021) was used to expand the nutrients, and as a case study to develop and test a semi-automated food matching process for fish and other aquatic products. The semi-automated food matching used the ISSCAAP groups (i.e. 50 groups in which commercial species are grouped based on their taxonomic, ecological and economic characteristics) and the harmonised food description which was based on the previous work of Grande and colleagues (2024), and coded in R.

The first step was the food entry classification/identification using the ASFIS list (FAO, 2022) which contains information of the scientific names and common names of 13,420 species for fisheries statistics and their corresponding taxonomic, ISSCAAP group and the 3-alpha code (i.e. a unique three letter code for each species allowing for easier inter-agency data exchange). When possible, food entries (i.e., fish species) were directly linked (i.e., joining two datasets together) using the scientific name. Where this did not successfully identify matches, an approach was developed to facilitate the semi-automatic matching of names, using approximate matching (e.g. fuzzy matching) based on the scientific name or the common names (in English).

The second step was the harmonisation of the food description based on the one-to-seven SUA item descriptor for fish and other aquatic products: "Fresh" (1), "Frozen Whole" (2), "Fillets" (3), "Frozen Fillets" (4), "Cured" (5), "Canned" (6) and "Preparations" (7). Whereby using string identification of terms such as "raw", "whole", "fillet" or "dried" each food entry was assigned to one of the seven food description groups.

Finally, food matching between the food entries in the Norwegian FCT (2021) and the SUA items for fish and other aquatic products were performed using the ISSCAAP group and the food description codes. For example, the food entry "Whiting, raw" was matched to SUA item demersal, fresh, whole (1514) according to the ISSCAAP group (32) and the food description code (1).

Manual identification of the remaining unmatched food entries and checks for the coherence of all the matches were performed. Food match quality criteria adapted from the "FAO/INFOODS Guidelines for food matching" (FAO/INFOODS, 2012b) was assigned to all matches, based on the similarity of the food description between the foods matched (Grande et al., 2024).

2.4.2. *Dealing with missing values*

When compiling an NCT, missing values should be avoided in the food components of interest and in those that are needed to (re-)calculate other nutrients, for example, retinol and provitamin A carotenoids to calculate vitamin A equivalents (Moltedo et al., 2021). Within this framework, scripts were developed to identify any missing values for individual food components, and to perform conventional and alternative approaches to address and reduce the number of missing values. These operations, which are described in the following sections, may affect the data quality and the derived results, hence, all of them are performed by independent functions and/or scripts that can be omitted if new and/or more accurate data become available. In addition, metadata were added (to the "*comments*" variable) for their identification, and for performing sensitivity analyses.

2.4.2.1. Food component imputation. Data imputation is used for food components in the following situations: 1) "borrowing" value(s) from similar food(s) in case of a missing value; 2) using a value reported in the original FCT when it should be calculated for harmonisation purposes (see Section 2.4.3) (for instance, beta-carotene equivalents should be calculated using values of provitamin A carotenoids; however, in cases when the individual components are not provided the beta-carotene equivalent value is imputed from the original FCT for the same food item); and 3) assuming zero, for example when a value is calculated and yielded a negative result, e.g., values for carbohydrates calculated by difference from the other proximate values, and when components are not naturally present in a food, including alcohol assumed to be zero in all foods except alcoholic drinks and some fermented products, and fibre in animal-source foods with exception of insects and food products and preparations (FAO/INFOODS, 2012a).

2.4.2.2. Food component combination. Certain food components are expressed in FCTs using different tagnames according to the fraction analysed (e.g., vitamin D3, vitamin D2, resulting in tagnames <ERGCAL> and <CHOCAL>, respectively), or the method of analysis (e.g., total vitamin B6 analysed with microbiological assay or HPLC, resulting in tagnames <VITB6A> and <VITB6C>, respectively). Here, a function combines and stores them into a new variable, where appropriate, which is named with the respective tagname and the word "*compiled*" (e.g., <VITB6-compiled >) and information about the original tagname is stored as metadata.

2.4.2.3. Food component back-calculation. There are cases where back-calculation (i.e., inferring or calculating one nutrient from other(s)) is needed, for example when calculating edible portion factor from refuse factor (Suppl. Mat. Eq.1a-b). Additionally, there are special instances where this is used for reducing the total number of missing values. For example, for retinol, or beta-carotene equivalent, when it could not be re-calculated (i.e., individual carotenoids values were missing) or imputed from beta-carotene equivalent, then in some specific foods beta-carotene equivalent or retinol can be back-calculated (Suppl. Mat. Eq. 3 and 4a-b) using an iteration of the equation to calculate vitamin A expressed as Retinol Activity Equivalent (RAE) and/or vitamin A as Retinol Equivalent (RE) (Suppl. Mat. Eq.2b-c).

2.4.3. *Food component re-calculation*

According to the FAO/INFOODS guidelines, some food components should be re-calculated from other food components even when they are reported in the original FCT. For example, sum of proximate (g), energy (kcal, kJ), carbohydrates available, by difference (g) among others (Suppl. Mat. Eq. 2a-g) (FAO/INFOODS, 2012a; Grande et al., 2024; Moltedo et al., 2021). Functions to perform those calculations were developed and combined into the R package: *NutritionTools* (Codd, Segovia de la Revilla, 2023).

2.5. *Quality checks and visualisation*

Iterative quality checking and processing script updating was used to ensure that any inconsistencies missed in the processing scripts were identified and rectified at the appropriate location in the processing steps. Identifying data inconsistencies that can lead to missing values is one of the main tasks when compiling and quality checking an FCT/NCT. Here, we used the *naniar package* (Tierney and Cook, 2023) for missing value visualisation and analysis. This step was essential to identify any typos or issues in previous steps, e.g., renaming the food components, unit of measurement transformations, and to evaluate the potential food components for inclusion/ exclusion for the final NCT depending on the number of missing values. Similarly, histograms, density plots, and boxplots were employed to assess the data availability, variability and identify potential implausible values. For instance, a script is available to generate a histogram of the sum of proximate components that allow for the identification of values within and outside the acceptable (95–105 g per 100 g of fresh weight, edible portion) and preferred (97–103 g per 100 g of fresh weight, edible portion) range or to identify items with unrealistic values (e.g., above 30 g of protein in fresh fish) as recommended by the FAO/ INFOODS guidelines (FAO/INFOODS, 2012a).

2.6. Data compilation and documentation

The final dataset corresponding to the NCT contained information on food components for fish and aquatic products as part of the Global NCT for SUA (Grande et al., 2024). Of which, energy and 15 nutrients were previously compiled by experts from FAO's Food and Nutrition Division, serving as a validation of the framework. To facilitate the comparison a script with a specific formatting structure was developed to generate and export the NCT into commonly used and understood formats, such as a Microsoft Excel workbook. Additionally, nine nutrients (total saturated, monounsaturated, polyunsaturated fatty acids, docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), vitamin B6, B12, copper, and selenium) were compiled and expanded for the fish and aquatic products to show-case the implementation of the framework. The complete NCT could be used within R or exported to other software, which can then be visually inspected, peer-reviewed and/or used in multiple software commonly available to researchers working with FCT/NCT datasets.

3. Results

Twelve FCTs of national, regional or global coverage from 2011–2021 were standardised, which included renaming food components and assigning tagnames, standardising units and denominators, and compiling food components and formatting into a unique structure to build the food composition data library (n=24,429 food entries). The original FCTs were in multiple data structures which conditioned the length and complexity of the scripts developed and used. For instance, well-formatted tabular FCTs, such as the FAO/INFOODS FCT for Western Africa (2019) or the Kenya FCT (2018), only needed one or two lines of code to import the dataset, while a Microsoft Access relational database, like US Department of Agriculture (USDA) (2019), may need to load the index of files, then identify the tables that are related to food composition data and then import those files. Table 1 presents the FCTs included and related information, including number of food items, and data provenance.

The classification of all fish and other aquatic products available in the selected FCTs (n=1846) was harmonised using the ASFIS list, i.e. identified according to their ISSCAAP group and assigned 3-alpha codes, and then were matched to one or more of the 95 SUA items for fish and other aquatic products. This resulted in a total of 4855 matched foods from the food composition library meaning that in many cases the same food from the FCTs was matched to more than one SUA item. One important source of duplication was the use of the same food items from FCTs for both "fresh" and "frozen" SUA items, since this description was rarely included in FCTs. Suppl. Table 3 shows the number of unique food items included per FCT whereas the Fig. 3 shows the number of effective food matched to each SUA item. Out of the total matched foods from the food composition library, 233 items were matched from the Norwegian FCT (2021), and 97 % (n= 225) were successfully matched using the semi-automatic matching developed here. When accounting for duplicates, on average, each SUA item was matched to 51 (range 1–290) foods from compiled FCTs, with the largest proportion extracted from The Standard Tables of Food Composition in Japan (2015) (19 %, n=947), followed by FAO/INFOODS Global FCT for Fish and Shellfish (uFiSh) (2016) (17 %, n=819) and US Department of Agriculture (USDA) (2019) (17 %, n=816), while the FCTs with fewest matches (<3 % of total) were the Brazilian FCT (2011) and the Kenya FCT (2018) (n=110 and n=118 respectively; Fig. 3).

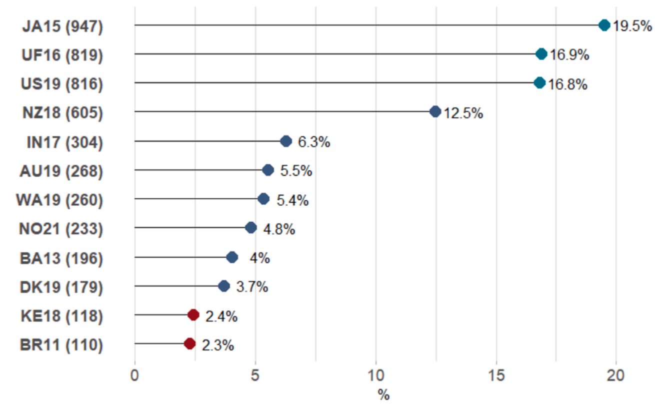


Fig. 3. The percentage of fish and other aquatic products contribution to the Global Nutrient Conversion Table from each Food Composition Table and Database, and in parenthesis is the number of food entries (including duplicated food entries being matched to multiple SUA items (see results section)) that were included. In alphabetical order: Australian Food Composition Database (AU19); Food Composition Table for Bangladesh (BA13); Brazilian FCT (TACO) (BR11); Frida: Food Database (DK19); Indian Food Composition Tables (IN17); Standard Tables of Food Composition in Japan (JA15); Kenya Food Composition Tables (KE18); The Norwegian Food Composition Table (NO21); New Zealand Food Composition Database (NZ18); uFiSh: FAO/INFOODS User Database for Fish and Shellfish (UF16); USDA National Nutrient Database for Standard Reference, Legacy Release (US19); FAO/INFOODS Food Composition Table for Western Africa (WA19).

Missing values were evaluated in the matched foods only (Fig. 4a) for each food component, and specific approaches were taken to resolve them, for instance, data for different tagnames were combined for certain components, using the function *nutri_combiner()* (Codd and Segovia de la Revilla, 2023). As a result, missing values for thiamin, fat and vitamin B6 were reduced by 99 %, 91 % and 88 %, respectively (Fig. 4b). Similarly, while most of the beta-carotene equivalent values were recalculated from the pro-vitamin A carotenoids, 6 % (n=289) were imputed from beta-carotene equivalent as presented in the original FCTs using the function *CARTBEQ_standardised()* (Codd and Segovia de la Revilla, 2023). After missing values were resolved, seven food components were re-calculated for matched foods using functions, for instance, sum of proximate, and energy (kcal)/(kJ), carbohydrates available, by difference, or vitamin A (RE)/(RAE) amongst others presented in Section 2.4.

General quality checks were then performed, first for the matched foods and then, for the average values corresponding to each of the 95 SUA items. For instance, 12 food entries were outside the acceptable range of the sum of proximate components (95—105 g), all of which were due to overestimation of the proximate values (Fig. 5a). When aggregated at SUA item level, the sum of proximate values, which were calculated based on the averaged/ re-calculated values, were all within acceptable range (FAO/INFOODS, 2012b; Greenfield and Southgate, 2003). In addition, protein was below 30 g per 100 g of fresh weight, edible portion for all the food entries considered "fresh fish" (e.g., "fresh", "frozen whole", "fillets", etc.) (Fig. 5b).

Finally, from the initial 89 tagnames collected from FCTs, data for 32 food components were compiled and reported for the 95 SUA items for fish and aquatic products (Supplementary Table 1A). From the food components compiled, energy, edible portion and other 21 food components, which were previously compiled by experts from FAO's Food and Nutrition Division, showed comparable results serving this compilation as a validation of the process, additionally nine nutrients (total saturated, monounsaturated, polyunsaturated fatty acids, docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), vitamin B6, B12, copper, and selenium) were compiled only for the fish and aquatic products and expanded as part of the Global NCT for SUA (Grande et al., 2024).

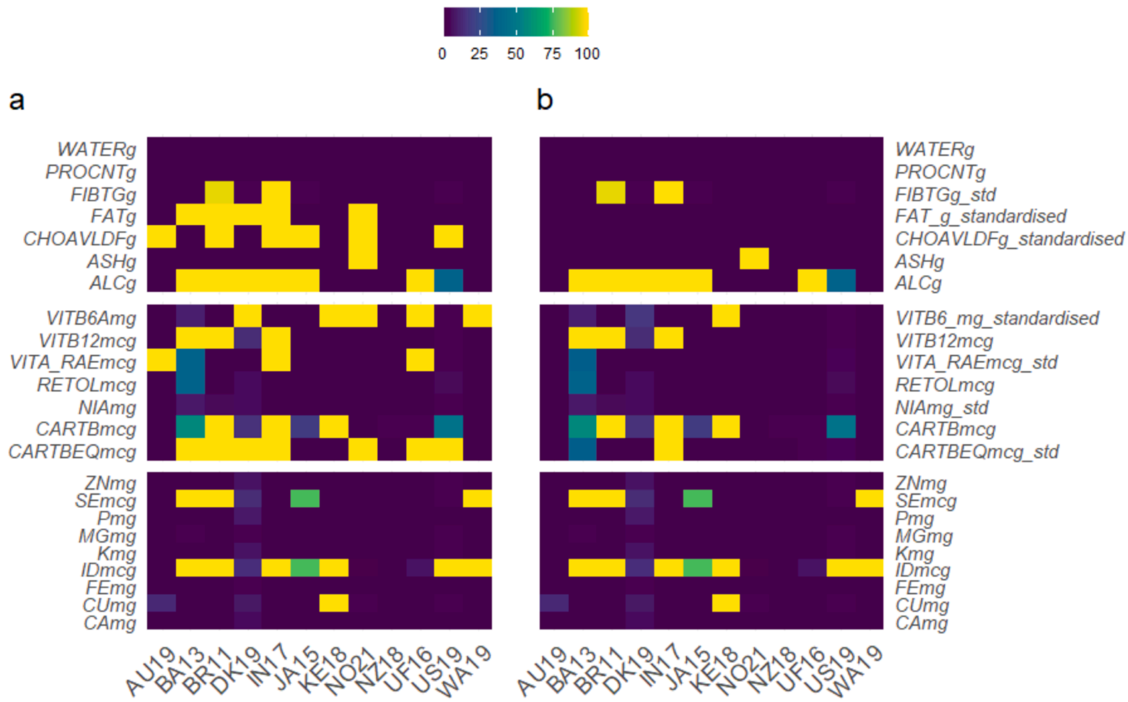


Fig. 4. Heatmap representing the missing values (% from 0 (dark blue) to 100 (yellow)) in proximate, vitamin and mineral values in the fish and other aquatic products after the standardisation and before dealing with the missing values (a), and after dealing with missing values (b) as part of the harmonisation steps.

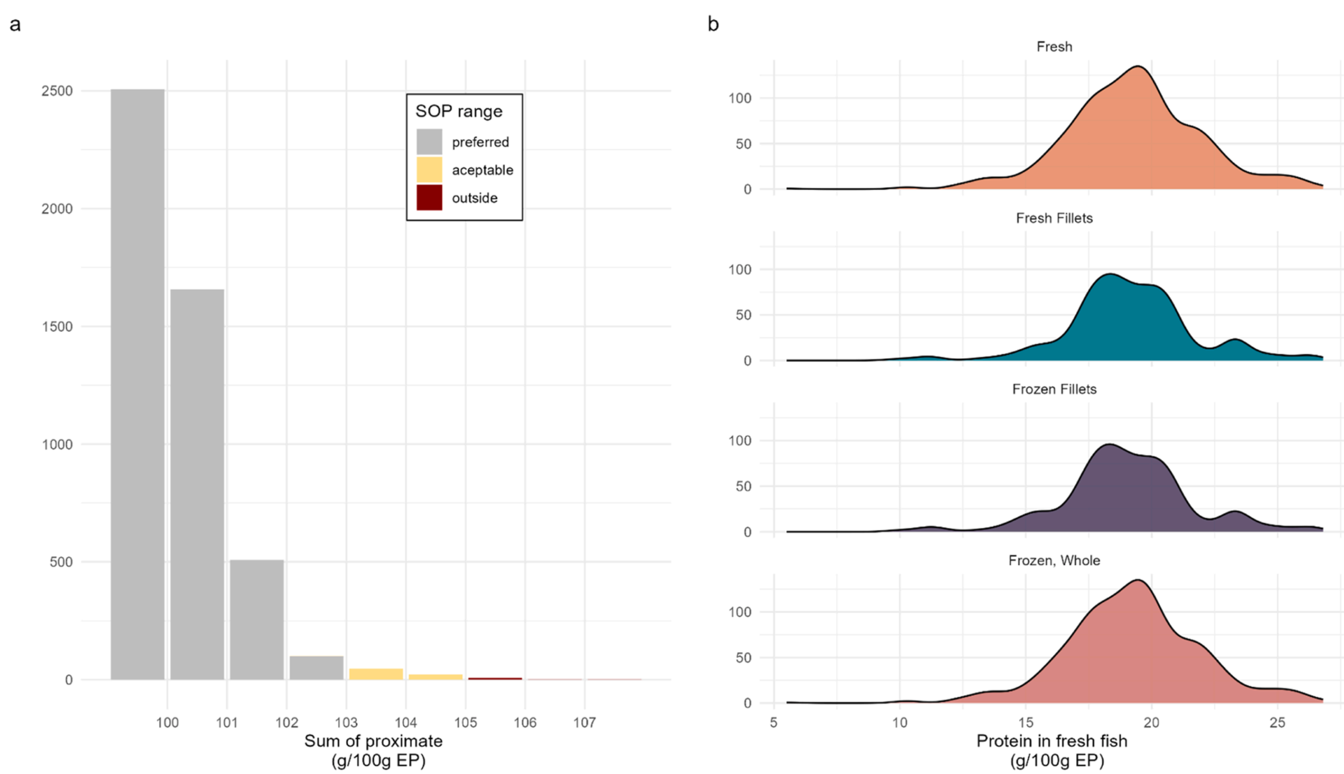


Fig. 5. Example visualisation for quality check of food composition or nutrient conversion tables: a) the histogram of the Sum of Proximate (SOP) with the grey bins representing the preferred range (97–103 g/ 100 g EP), the yellow bins representing outside preferred range and within the acceptable range (93–95 and 103–105 g/ 100 g EP) and the red bins representing the outside acceptable range, and b) the density plots of the protein (g/ 100 g EP) content in fresh fish entries grouped by their state (e.g., raw, frozen, fillet etc).

This global NCT was applied to the SUA items and statistics on energy and nutrient availability at national level and are presented for 186 countries as part of the FAOSTAT Food and Diet Domain (FAO, 2024a), a web-hosted portal dedicated to the dissemination of statistics on different types of dietary data.

The R scripts developed in this study are presented in the public repository published in GitHub (https://github.com/LuciaSegovia/FAO-fisheries) which provides commands for cleaning and standardisation of the 12 FCTs, harmonisation and compilation of the NCT, including visualisation and quality checks, functions to re-calculate energy and other food components, food matching aid and FCT/NCT output formatting. All functions are published as an R package: *NutritionTools* (Codd and Segovia de la Revilla, 2023). Furthermore, a food composition data library with the nine open and freely accessible FCTs can be compiled by running the scripts in the repository. For the other FCTs of interest, scripts could be re-used if data access/licences permit, or scripts can be adapted to current publicly available version of the data (e.g. Food Standards Australia New Zealand, release 1 (2019), to script to Food Standards Australia New Zealand, release 2 (2023), see Table 1). Moreover, a template for processing new FCTs is included together with a manual that describes every step and provides guidance on standardisation decisions.

4. Discussion

4.1. Data standardisation and harmonisation framework for food composition data

A framework and the R tools to ingest, process and standardise FCTs and to compile and report FCTs/NCTs was developed. Our objective was to increase efficiency, reproducibility, and transparency in the processing of food composition data for nutrition.

Several projects have undertaken standardisation/harmonisation of FCTs for Europe, such as the European Prospective Investigation into Cancer and Nutrition (EPIC) Nutrient Database (ENDB) (Slimani et al., 2007), European Food Information Resource (EuroFIR) (Finglas et al., 2014) and, more recently Stance4Health (Hinojosa-Nogueira et al., 2021). All of them, including our project, have faced analogous data challenges, i.e., lack of food component and food description standardisation, diverse measurement of units, missing and/or implausible values, etc. Despite the amount of thought and effort invested in the cleaning and standardisation process from numerous experts and projects, most of the steps and/or clean datasets are not openly available. This results in a lack of transparency and reproducibility of the methods, concerns about the reusability of the cleaned FCTs/NCTs, and ultimately, researchers repeating the process (Clancy et al., 2015). To address these issues, we developed an open science workflow, generating and publishing all the data processing steps as scripts in a format compatible with an open and freely available software. Thus, our processing can be readily reproduced by anyone able to access the original FCTs, to avoid continued duplication of effort. Moreover, the use of scripted approaches allows for full audit of the data manipulation/decisions performed when generating a food composition library and NCT (Coates et al., 2017; Micha et al., 2018).

4.2. Framework re-usability: the users' journey

There are numerous ways in which the scripts and functions can be used as provided, or further enhanced and adapted, to support the principles of FAIR data in food composition science and nutritional assessments using NCTs. Four example users of the open science workflow are provided to demonstrate relevance for the nutrition research community (Fig. 1). Three of the user examples are described in detail below while the fourth, represented in the green band in Fig. 1, and which is interested in standardising a new FCT only, would just need to adapt the standardisation scripts, as described in the Example 2.

In all cases, the first step will be to visit the GitHub repository for the scripts and functions used in this study and follow the instructions to obtain the data and tools as needed. Data provenance are reported in Table 1 and within the repository.

4.2.1. Example user 1: reproducing the global NCT for SUA for the fish and other aquatic products

The first example represented in the white band in the Fig. 1 (as "Framework") presents a user aiming to reproduce the steps undertaken in this project, to replicate the energy and nutrient values of the 95 SUA items for fish and other aquatic products compiled as part of the global NCT (Grande et al., 2024), and for which statistics are presented in the Food and Diet Domain (FAO, 2024a). The annotated scripts can be followed to review decisions on data cleaning and standardisation. With the scripts, which contain the necessary functions, and FCTs all in place, the user could run the workflow and obtain the NCT with 95 SUA items for fish and other aquatic products in less than five minutes. The majority of the scripts will work even if one or more FCT is not included. However, we would recommend that our documented decisions are reviewed and consciously adopted, or adapted, by the user, according to their needs. Nonetheless, the workflow will save considerable time compared to the effort required to recreate all the steps involved; it will also generate a traceable record of decisions made within the NCT preparation, which is often lacking (Clancy et al., 2015; Coates et al., 2017; Pennington et al., 2007; Segovia de la Revilla et al., 2023).

4.2.2. Example user 2: standardising an additional FCT and expanding the current NCT

The second example presents the situation of a food composition data compiler who wishes to add an additional FCT to the current list, for instance, to update or expand the energy and/or nutrient for the 95 SUA items for fish and other aquatic products. These steps are presented in the blue band in Fig. 1 (as "NCT Compiler (Updating)") and require more time and technical expertise (i.e., R programming skills) from the user than required in the first example user, however by copying and adapting existing FCT import scripts, this activity can be accomplished with considerably reduced effort, and increased transparency compared to manual inclusion of the additional FCT.

Firstly, to standardise a new FCT, the user is advised to check the quality of the FCT by using the FAO/INFOODS evaluation framework (Charrondiere et al., 2023). Following this, a template script, which can be accessed here (https://github.com/LuciaSegovia/FAO-fisheries/tree/main/00_template) for import and standardisation is provided which is designed to help the user navigate the scripting of the steps 1 to 3 of the framework (Fig. 2). The template script provides guidance and options covering the most common tasks for an array of FCT formats, such as those detailed in Sections 2.2 and 2.3 of the methods. After completing the standardisation steps the user can either export the standardised FCT in a tabular format (e.g., Microsoft Excel, text-delimited), or include it in the food composition data library. Once an FCT is standardised no extra user inputs are needed for merging it with the food composition data library available in the repository (which can be obtained by running the *merging_all.R* script). The R scripts that perform the harmonisation steps (Fig. 2) (e.g., harmonisation of food names and description, dealing with missing values, etc.) should be reviewed and updated and the R scripts re-run. For instance, if the aim is to update energy and/or nutrient information for the 95 SUA items for fish and other aquatic products with this additional FCT, the user needs to adapt the scripts of the semi-automatic standardising of the food description and food matching between the food entries in the additional FCT and the 95 SUA items: this can be adapted from the *NO21_harmonising.R* script. Then, the other harmonisation, quality checks, visualisation and formatting steps can be performed as described in the method section using existing scripts, without further adaption, which produce the traceable updated/extended NCT and metadata.

4.2.3. Example user 3: The Nutrient Conversion Table compiler: Re-usability of the food composition data library

The third example represents a user wishing to obtain a survey specific NCT, e.g. for the list of foods reported as consumed and/or acquired in a household consumption and expenditure survey (as the "NCT Compiler" in the yellow band in Fig. 1). This user needs the highest R/data literacy of the examples provided whilst it requires similar, or less time and effort than in the second example. This is because all the import and standardisation tasks for the entirety of the FCTs, comprising steps 1 – 3 of Fig. 2, are already scripted and documented. The product, after obtaining the original FCTs and running the scripts, is a standardised food composition data library which contains food items (n=24,429) from 12 high quality FCTs covering different regions of the world. From the library the user can extract information from 32 food components and/or benefit from the harmonisation and formatting scripts available to compile a new NCT. This, in turn, would reduce considerably the time and effort needed while increasing the transparency and reproducibility of the output. Nevertheless, the user would need to implement and adapt some of the scripts/steps, such as, the food matching between the food composition data library and the food consumption dataset, setting rules and priorities as appropriate (e.g. using a country/region specific FCT as the main source for performing the food matching). An additional function was developed within this framework, the *Fuzzy_Matcher(),* that provides an aid to this time consuming step. The output, a user-led matched dataset, can be integrated with the subsequent functions and scripts to perform the rest of the harmonisation, quality check and formatting steps. Particular attention should be given to the food matches which should be carefully checked for coherence and context relevancy. An example of the use of food composition data library and the *Fuzzy_Matcher()* for generating household survey NCTs can be found in the MAPS project repository here: [https://github.com/micronutrientsupport].

4.3. Strengths and limitations

4.3.1. Strengths

This study generated a novel open science workflow that can increase the findability, interoperability and reproducibility of FCTs and derived NCTs. The workflow will reduce the time and effort needed by other food composition data compilers and users, as exemplified by three use-cases. By providing a framework and the R tools (i.e., repository) in a freely available and open software, an array of different users can adopt and adapt these steps into their own workflow.

Furthermore, the scripted approach outlined here ensures that each data processing decision and assumption that may influence outputs derived from FCTs, is recorded and annotated. The documentation (reporting and metadata) proposed within this framework (e.g., reusable scripts and functions, reporting and exporting structure, etc.) provides a solution to the insufficient reporting of the food composition data found in most of FCTs/NCTs (Bruyn et al., 2016; Ispirova et al., 2020; Segovia de la Revilla et al., 2023). Additionally, the detailed documentation and visualisation for quality checks can aid data processing decisions and allow for revisiting assumptions. This improves transparency and reproducibility, and reduces the uncertainty around data and derived outputs.

4.3.2. Limitations

One of the main limitations of the study is that, although all the scripts, functions and decisions made are recorded and publicly available, the raw data used (i.e., FCTs) are not all publicly available, and some have been updated since the publication of the global NCT. For instance, the Food Standards Australia New Zealand (FSANZ Food Standards Australia New Zealand, 2019) has been replaced with a newer version, and the Indian Food Composition Tables (2019) is only available as pdf. Furthermore, the impact of the data processing decisions (including imputation, calculation, etc.) on the energy and nutrient supplies was not evaluated. For instance, the influence of the combination of different tagnames reporting different methods of analysis (i.e., used for vitamin B6) or transforming trace and below detection limit values to zero had on mean nutrient values for each SUA item.

Another limitation related to the quality checks is that despite the sum of proximate components frequently being used an indicator of data quality, as suggested by most of the food composition compilation guidelines (FAO/INFOODS, 2012a; Greenfield and Southgate, 2003), it is only a reliable measurement of quality for analytical values. Here, we used carbohydrates available, by difference (i.e., calculated from the other proximate values) which is not a reliable component for its calculation given that the same proximate components are used in both the calculation of carbohydrates available, by difference and sum of proximate. However, this decision was made as carbohydrates available, by weight were not available in all FCTs used in the present work and for most of the SUA items included in this case study this component would be zero or assumed zero (FAO/INFOODS, 2012a). Other quality checks that could be used instead are 95 % confidence intervals and/or implausible values detection per food group. Nevertheless, these checks may not be effective for all food components as there is high variability in the concentration of certain food components due to multiple factors, such as: broad spectrum of food entries within each SUA item (e.g., sea urchins and turtles are considered under the same SUA item), often compounded by limited number of values available for certain food components and, low or uncertain quality of the values. This uncertainty around the food component values makes the establishment of quality checks difficult and increases the need for expert assessment and inputs (FAO, 2024a).

Finally, we acknowledge that resources are needed to regularly update the R scripts and packages to maintain its functionality and compatibility with newer R versions and packages, ensuring its relevance for the community. Further development of the framework and the tools could also be undertaken by the network of food composition experts and data users who will benefit from this repository.

5. Conclusion and recommendations

To the best of our knowledge, this is the first example of a comprehensive method to develop NCTs suitable for global application in a reproducible, reusable, efficient, and transparent manner. The NCT output has been used in the published statistics based on SUA data for 186 countries on the FAOSTAT Food and Diet Domain (FAO, 2024a; 2024b). Statistics comprise energy and nutrient availability for 26 nutrients including the nine nutrients expanded by the present work for fish and other aquatic products.

Open science offers opportunities to greatly reduce the resources required to compile food composition data for nutrition research. Here, scripts and functions are provided aimed at making the data processing more reproducible, reusable, efficient, and transparent. To further support open research, the food composition community should agree and implement standardised practices for data management and documentation (i.e., minimum information standards, metadata, ontologies) that improve reproducibility, transparency, efficiency and (re-) usability of future FCTs and NCTs.

Disclaimer

The views expressed in this publication are those of the author(s) and do not necessarily reflect the views or policies of the Food and Agriculture Organization of the United Nations.

CRediT authorship contribution statement

E. Louise Ander: Writing – review & editing, Writing – original draft, Project administration, Investigation, Funding acquisition, Conceptualization. **Lucia Segovia de la Revilla:** Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Thomas Codd:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation. **Edward J.M.**

Joy: Writing – review & editing, Supervision, Project administration, Methodology. **Liberty Mlambo:** Writing – review & editing, Validation, Software. **Fernanda Grande:** Writing – review & editing, Supervision, Resources, Methodology. **Doris Rittenschober:** Writing – review & editing, Resources, Methodology. **Ana Moltedo:** Writing – review & editing, Supervision, Resources, Methodology. **Bridget A. Holmes:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Food and Agriculture Organization of the United Nations (FAO) (PO Number 352277). In addition, this work was supported in part by the Bill & Melinda Gates Foundation [INV-002588; Micronutrient Action Policy Support (MAPS) project]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. The funder had no role in the study design, data analysis, decision to publish, or preparation of the manuscript. We also would like to thank Claire Dooley and Elaine L. Ferguson for their help on the earliest versions of the manuscript

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfca.2024.106894.

Data Availability

The authors do not have permission to share data.

References

Andrés-Hernández, L., Blumberg, K., Walls, R.L., Dooley, D., Mauleon, R., Lange, M., Weber, M., Chan, L., Malik, A., Møller, A., Ireland, J., Segovia, L., Zhang, X., Burton-Freeman, B., Magelli, P., Schriever, A., Forester, S.M., Liu, L., King, G.J., 2022. Establishing a common nutritional vocabulary - from food production to diet. Front. Nutr. 9.

Bruyn, J., de, Ferguson, E., Allman-Farinelli, M., Darnton-Hill, I., Maulaga, W., Msuya, J., Alders, R., de Bruyn, J., Ferguson, E., Allman-Farinelli, M., Darnton-Hill, I., Maulaga, W., Msuya, J., Alders, R., 2016. Food composition tables in resource-poor settings: exploring current limitations and opportunities, with a focus on animal-source foods in sub-Saharan Africa. Br. J. Nutr. 116, 1709–1719. https://doi.org/10.1017/S0007114516003706.

Chan, L., Vasilevsky, N., Thessen, A., McMurry, J., Haendel, M., 2021. The landscape of nutri-informatics: a review of current resources and challenges for integrative nutrition research. Database 2021. https://doi.org/10.1093/database/baab003.

Charrondiere, U.R., Stadlmayr, B., Grande, F., Vincent, A., Oseredczuk, M., Sivakumaran, S., Puwastien, P., Judprasong, K., Haytowitz, D., Gnagnarella, P., 2023. FAO/INFOODS Evaluation Framework to Assess the Quality of Published Food Composition Tables and Databases: User Guide. FAO, Rome, Italy. https://doi.org/10.4060/cc5371en.

Clancy, A.K., Woods, K., McMahon, A., Probst, Y., 2015. Food composition database format and structure: a user focused approach. PLOS ONE 10, e0142137. https://doi.org/10.1371/journal.pone.0142137.

Coates, J.C., Colaiezzi, B.A., Bell, W., Charrondiere, U.R., Leclercq, C., 2017. overcoming dietary assessment challenges in low-income countries: technological solutions proposed by the international dietary data expansion (INDDEX) project. Nutrients 9.

Codd, T., Segovia de la Revilla, 2023. NutritionTools: Tools for the Organisation, Matching, Calculation, and Summarisation of Nutrition Data.

Dooley, D.M., Griffiths, E.J., Gosal, G.S., Buttigieg, P.L., Hoehndorf, R., Lange, M.C., Schriml, L.M., Brinkman, F.S.L., Hsiao, W.W.L., 2018. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. Npj Sci. Food 2, 23. https://doi.org/10.1038/s41538-018-0032-6.

DTU (Technical University of Denmark) Food Institute, 2019. Frida Food Databank. Version 4. Copenhagen, DTU, National Food Institute. [Cited 29 August 2023]. https://frida.fooddata.dk.

Durazzo, A., Astley, S., Kapsokefalou, M., Costa, H.S., Mantur-Vierendeel, A., Pijls, L., Bucchini, L., Glibetić, M., Presser, K., Finglas, P., 2022. Food composition data and tools online and their use in research and policy: EuroFIR AISBL contribution in 2022. Nutrients 14. https://doi.org/10.3390/nu14224788.

Durazzo, A., Lucarini, M., 2022. Editorial: databases and nutrition. Front. Nutr. 9. https://doi.org/10.3389/fnut.2022.853600.

Ene-Obong, H., Schönfeldt, H.C., Campaore, E., Kimani, A., Mwaisaka, R., Vincent, A., Ati, J.E., Kouebou, P., Presser, K., Finglas, P., Charrondiere, U.R., 2019. Importance and use of reliable food composition data generation by nutrition/dietetic professionals towards solving Africa's nutrition problem: constraints and the role of FAO/INFOODS/AFROFOODS and other stakeholders in future initiatives. Proc. Nutr. Soc. 78, 496–505. https://doi.org/10.1017/S0029665118002926.

European Food Safety Authority, 2015. The Food Classification and Description System FoodEx 2 (revision 2) (No. 804), EFSA supporting publication.

FAO, 2016. FAO/INFOODS Global Food Composition Database for Fish and Shellfish. Version 1.0 – uFiSh1.0. Rome.

FAO, 2021. FAO Yearbook. Fishery and Aquaculture Statistics 2019/FAO annuaire. Statistiques des pêches et de l'aquaculture 2019/FAO anuario. Estadísticas de pesca y acuicultura 2019. FAO. https://doi.org/10.4060/cb7874t.

FAO, 2022. ASFIS List of Species for Fishery Statistics Purposes. Fisheries and Aquaculture Division [WWW Document]. URL. https://www.fao.org/fishery/en/collection/asfis/en (accessed 2.21.23).

FAO, 2024a. Food and Diet. Statistics on Dietary Data (No. 82), FAOSTAT Analytical Briefs. Rome. https://openknowledge.fao.org/handle/20.500.14283/cc9454en.

FAO, 2024b. Food and Diet / Availability (Supply utilization accounts). FAOSTAT. https://www.fao.org/faostat/en/#data/SUA.

FAO & Government of Kenya, 2018. Kenya Food Composition Tables [online]. Nairobi. https://www.fao.org/infoods/infoods/standards-guidelines/food-component-identifiers-tagnames/en/.

FAO/INFOODS, 2012a. FAO/INFOODS Guidelines for Checking Food Composition Data prior to the Publication of a User Table/Database-Version 1.0. FAO, Rome.

FAO/INFOODS, 2012b. FAO / INFOODS Guidelines for Food Matching Version 1.2.

FAO/INFOODS, 2012c. FAO/ INFOODS Guidelines for Converting Units, Denominators and Expressions Version 1.0. FAO/ INFOODS.

Finglas, P.M., Berry, R., Astley, S., 2014. Assessing and improving the quality of food composition databases for nutrition and health applications in Europe: the contribution of EuroFIR123. Adv. Nutr. 5, 608S–614S. https://doi.org/10.3945/an.113.005470.

FSANZ (Food Standards Australia New Zealand), 2019. Australian Food Composition Database. Release 1. Canberra. http://www.foodstandards.gov.au.

Grande, F., Ueda, Y., Masangwi, S., Holmes, B., 2024. Global Nutrient Conversion Table for FAO Supply Utilization Accounts. FAO, Rome, Italy. https://doi.org/10.4060/cc9678en.

Greenfield, H., Southgate, D.A.T., 2003. Food Composition Data: Production, Management, and Use. FAO, Rome.

Hinojosa-Nogueira, D., Pérez-Burillo, S., Navajas-Porras, B., Ortiz-Viso, B., de la Cueva, S.P., Lauria, F., Fatouros, A., Priftis, K.N., González-Vigil, V., Rufián-Henares, J.Á., 2021. Development of an unified food composition database for the European project "Stance4Health". Nutrients 13, 4206. https://doi.org/10.3390/nu13124206.

Ispirova, G., Eftimov, T., Koroušić, B., 2017. Mapping Food Composition Data from Various Data Sources to a Domain-Specific Ontology 8. https://doi.org/10.5220/0006504302030210.

Ispirova, G., Eftimov, T., Seljak, B.K., 2020. Evaluating missing value imputation methods for food composition databases. Food Chem. Toxicol. 141, 111368. https://doi.org/10.1016/j.fct.2020.111368.

Joy, E.J.M., Kumssa, D.B., 2022. Nutrient accounting in global food systems, 678–678 Nat. Food 3. https://doi.org/10.1038/s43016-022-00593-w.

Kapsokefalou, M., Roe, M., Turrini, A., Costa, H.S., Martinez-Victoria, E., Marletta, L., Berry, R., Finglas, P., 2019. Food composition at present: new challenges. Nutrients 11, 1714. https://doi.org/10.3390/nu11081714.

Klensin, J.C., Feskanich, D., Lin, V., Truswell, A.S., Southgate, D.A.T., 1989. Identification of food components for INFOODS data interchange. The United Nations University.

Longvah, T., Ananthan, R., Bhaskarachary, K. & Venkaiah, K., 2017. Indian Food Composition Tables. Hyderabad, India, National Institute of Nutrition, Indian Council of Medical Research.

Luo, H., Dodd, K.W., Arnold, C.D., Engle-Stone, R., 2021. Introduction to the simple macro, a tool to increase the accessibility of 24-hour dietary recall analysis and modeling. J. Nutr. https://doi.org/10.1093/jn/nxaa440.

MEXT (Ministry of Education, Culture, Sports, Science and Technology), 2015. Standard Tables of Food Composition in Japan 2015. 7th revised edition. Tokyo, Official Gazette Co-operation of Japan. https://www.mext.go.jp/en/policy/science_technology/policy/ title01/detail01/1374030.htm.

Micha, R., Coates, J., Leclercq, C., Charrondiere, U.R., Mozaffarian, D., 2018. Global dietary surveillance: data gaps and challenges. Food Nutr. Bull. 39, 175–205. https://doi.org/10.1177/0379572117752986.

Møller, A., Ireland, J., 2018. LanguaL™ 2017 - Thesaurus. https://doi.org/10.13140/RG.2.2.23131.26404.

Moltedo, A., Álvarez-Sánchez, C., Grande, F., Charrondiere, U.R., 2021. The complexity of producing and interpreting dietary vitamin A statistics. J. Food Compos. Anal. 100, 103926. https://doi.org/10.1016/j.jfca.2021.103926.

Moshfegh, A.J., Rhodes, D.G., Martin, C.L., 2022. National food intake assessment: technologies to advance traditional methods. Annu. Rev. Nutr. 42, 401–422. https://doi.org/10.1146/annurev-nutr-062320-110636.

NEPA-UNICAMP (Núcleo de Estudos e Pesquisas em Alimentação – Universidade Estadual de Campinas), 2011. Tabela Brasileira de Composição de Alimentos [Brazilian Food Composition Table] (TACO). 4th edition. Campinas, Brazil. https://www.nepa.unicamp.br/publicacoes/tabela-taco-pdf.
New Zealand Institute for Plant and Food Research Limited & Ministry of Health. 2019. New Zealand FOODfiles 2018 Manual. New Zealand Food Composition Database. Auckland. https://www.foodcomposition.co.nz/foodfiles.
Norwegian Food Safety Authority, 2021. Norwegian Food Composition Database. Oslo. https://www.matvaretabellen.no.
Pennington, J.A.T., Stumbo, P.J., Murphy, S.P., McNutt, S.W., Eldridge, A.L., McCabe-Sellers, B.J., Chenard, C.A., 2007. Food composition data: the foundation of dietetic practice and research. J. Am. Diet. Assoc. 107, 2105–2113. https://doi.org/10.1016/j.jada.2007.09.004.
Posit team, 2023. RStudio: Integrated Development for R.
R Core Team, 2023. R: A language and environment for statistical computing.
Rittenschober, D., Stadlmayr, B., Nowak, V., Du, J., Charrondiere, U.R., 2016. Report on the development of the FAO/INFOODS user database for fish and shellfish (uFiSh) – Challenges and possible solutions. Food Chem. 193, 112–120. https://doi.org/10.1016/j.foodchem.2014.08.055.
Segovia de la Revilla, L., Ferguson, E.L., Dooley, C., Osman, G., Ander, E.L., Joy, E.J.M., 2023. The availability and geographic location of open-source food composition data used to estimate micronutrient intakes in sub-Saharan Africa: a scoping review. J. Food Compos. Anal. 120, 105322. https://doi.org/10.1016/j.jfca.2023.105322.
Shaheen N., Rahim A.T.M.A., Mohiduzzaman Md, Banu C.P., Bari Md L., Tukun A.B., Mannan M.A., Bhattacharjee L., Stadlmayr B., 2013. Food Composition Table for Bangladesh. Institute of Nutrition and Food Science, Centre for Advanced Research in Sciences, University of Dhaka. https://www.fao.org/infoods/infoods/tables-and-databases/asia.
Slimani, N., Deharveng, G., Unwin, I., Southgate, D. a T., Vignat, J., Skeie, G., Salvini, S., Parpinel, M., Møller, A., Ireland, J., Becker, W., Farran, A., Westenbrink, S., Vasilopoulou, E., Unwin, J., Borgejordet, Å., Rohrmann, S., Church, S., Gnagnarella, P., Casagrande, C., van Bakel, M., Niravong, M., Boutron-Ruault, M.C., Stripp, C., Tjønneland, A., Trichopoulou, A., Georga, K., Nilsson, S., Mattisson, I., Ray, J., Boeing, H., Ocké, M., Peeters, P.H.M., Jakszyn, P., Amiano, P., Engeset, D., Lund, E., Santucci de Magistris, M., Sacerdote, C., Welch, A., Bingham, S., Subar, A. F., Riboli, E., 2007. The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. Eur. J. Clin. Nutr. 61, 1037–1056. https://doi.org/10.1038/sj.ejcn.1602679.
Tierney, N., Cook, D., 2023. Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. J. Stat. Softw. 105, 1–31. https://doi.org/10.18637/jss.v105.i07.
Traka, M.H., Plumb, J., Berry, R., Pinchen, H., Finglas, P.M., 2020. Maintaining and updating food composition datasets for multiple users and novel technologies: current challenges from a UK perspective. Nutr. Bull. 45, 230–240. https://doi.org/10.1111/nbu.12433.
United States Department of Agriculture (USDA), 2019. Agricultural Research Service, Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Legacy. Version Current: April 2019. Washington, DC, Agricultural Research Service, Nutrient Data Laboratory. https://fdc.nal.usda.gov/ download-datasets.html.
Vincent, A., Grande, F., Compaoré, E., Amponsah Annor, G., Addy, P.A., Aburime, L.C., Ahmed, D., Bih Loh, A.M., Dahdouh Cabia, S., Deflache, N., Dembélé, F.M., Dieudonné, B., Edwige, O.B., Ene-Obong, H.N., Fanou Fogny, N., Ferreira, M., Omaghomi Jemide, J., Kouebou, P.C., Muller, C., Nájera Espinosa, S., Ouattara, F., Rittenschober, D., Schönfeldt, H., Stadlmayr, B., van Deventer, M., Razikou Yiagnigni, A. & Charrondière, U.R., 2020. FAO/INFOODS Food Composition Table for Western Africa (2019) User Guide & Condensed Food Composition Table / Table de composition des aliments FAO/INFOODS pour l'Afrique de l'Ouest (2019) Guide d'utilisation & table de composition des aliments condensée. Rome, FAO. http://www.fao.org/infoods/infoods/tables-and-databases/faoinfoods-databases.
Zeb, A., Soininen, J.-P., Sozer, N., 2021. Data harmonisation as a key to enable digitalisation of the food sector: a review. Food Bioprod. Process. 127, 360–370. https://doi.org/10.1016/j.fbp.2021.02.005.

5 Chapter 5: Spatial food composition models for use in dietary micronutrient assessment: a case study of maize grain and selenium status in Malawi

This chapter presents the evaluation of 10 level of spatial aggregation of georeferenced maize selenium composition collected in Malawi to estimate dietary selenium intakes, and the comparison with georeferenced plasma Se concentration in women (15-49 years old) in the same country to identify the level of aggregation of Se concentration in maize suitable for dietary assessment.

The repository with all the study protocol, data exploration and all the code to replicate the analysis are available here.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	2006000	**Title**	Ms
First Name(s)	Lucia		
Surname/Family Name	Segovia de la Revilla		
Thesis Title	Geospatial analysis of food composition data to estimate population micronutrient intake		
Primary Supervisor	Edward Joy		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	International Journal of Health Geographics
Please list the paper's authors in the intended authorship order:	Lucia Segovia de la Revilla, Claire Dooley, Elaine L. Ferguson, R. Murray Lark, Hakunawadi A. Pswarayi, Gareth Osman, Alexander A. Kalimbira, Patson C. Nalivata, E. Louise Ander, Martin R. Broadley, Edward J.M. Joy
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conceptualised, performed the analysis and was the primary writer. Claire Dooley, R. Murray Lark and Hakunawadi A. Pswarayi helped reviewing and conceptialisation of the methods. R. Murray Lark, Hakunawadi A. Pswarayi, Gareth Osman, Alexander A. Kalimbira, Patson C. Nalivata, E. Louise Ander, Martin R. Broadley helped reviewing the methodology and the manuscript. Claire Dooley, Elaine L. Ferguson and Edward J.M. Joy helped with the conceptialisation and reviewing the manuscript.

SECTION E

Student Signature	Lucia Segovia de la Revilla
Date	10/12/2024

Supervisor Signature	EJMJ - Edward Joy
Date	06/01/2025

Spatial food composition models for use in dietary micronutrient assessment: a case study of maize grain and selenium status in Malawi

Abstract

Soil characteristics are a major driver of mineral micronutrient content in crops, which, in contexts of localised food systems, can greatly impact the population dietary intakes and status of those mineral micronutrients. In such contexts, the use of georeferenced food composition data may be necessary to accurately estimate population risk of micronutrient deficiency, yet this is not common practice. This study evaluated different levels of spatial aggregation for high resolution maize grain Se concentration in Malawi aiming to identify a suitable spatial (area level) aggregation of georeferenced staple grain concentration data for estimating dietary selenium intakes. We use the case study of maize grain concentration and selenium status in Malawi, where previous studies of blood and urinary biomarkers revealed marked spatial variability in selenium status, despite relatively homogeneous food patterns across the country, and where maize provides >60% of dietary energy and selenium intakes.

Georeferenced biomarker and socio-economic data for Malawian women (n=745) were obtained from the 2015-16 Malawian DHS-Micronutrient survey and combined with different subnational area-level aggregations of maize grain selenium concentration and potential confounders. Ten Bayesian spatial models, using Integrated Nested Laplace Approximation (INLA) for each area-level aggregation (10-60km buffers, and enumeration area (EA) group and district level), were tested and compared using the model performance metric, Deviation Information Criterion (DIC).

Maize grain selenium concentration was the main explanatory variable for plasma selenium concentration, followed by age, irrespective of the maize aggregation area-level. Based on the DIC, all models performed similarly with the smaller area-level aggregation of maize grain selenium concentration (<30 km) performing better than larger area-levels (>30 km). Our findings demonstrate the influence of high spatial resolution crop selenium concentration data in Se status and its potential for generating accurate estimates of risks for selenium inadequacy at sub-national scales in Malawi. Our method can be applied for other micronutrients and other settings, where food systems are localised.

Keywords: Biofortification; Biomarkers; Food Composition Data; Food System; GeoNutrition; Maize; Mineral micronutrients; Selenium; Spatial modelling

Background

Micronutrient deficiencies are widespread globally with varying prevalence rates between and within countries. Subnational estimates of dietary micronutrient intakes are required to inform the design, delivery, monitoring and evaluation of programmes to improve nutrition, like large-scale food fortification (Neufeld et al., 2017; Popkin et al., 2020). Heterogeneity of dietary micronutrient intakes at subnational scale may relate to a number of factors including urban/rural residency, socioeconomic status, and, particularly in the context of localised food systems, spatial variation in crop micronutrient content. A growing number of studies have explored and reported spatial variation in crop mineral concentrations within countries. For example, it is likely to be important for dietary selenium (Se) intakes in Malawi, Ethiopia and other countries (Ahmad et al., 2021; Gashu et al., 2021, 2019; Ligowe et al., 2020; Long et al., 2019, Watts et al., 2019). These studies have shown that concentration of Se in crops and livestock feed is highly variable and is strongly influenced by the concentration of the plant-available Se in the soil and other environmental factors. These studies suggested that using national averages of the Se composition in foods may not accurately estimate the Se intake of populations living in different areas of a country.

Despite the relevance of the spatial variability for dietary Se intake and deficiency risks, the use of georeferenced crop Se concentration data (i.e. adding geographic information to each sample or data point) and the use of geostatistical methods (i.e. incorporating spatial information to analysis) are not common practice in nutritional assessment (Ene-Obong et al., 2019; Segovia de la Revilla et al., 2023). The Se intakes of populations are often evaluated using food composition data sourced from national scale published food composition tables and databases (FCTs). However, to date, FCTs do not provide spatially disaggregated information of food components, despite recognition of the importance of spatial variability of food components and advocacy for the development of national and sub-national FCTs for over 20 years (Ene-Obong et al., 2019; Grande et al., 2016; Greenfield & Southgate, 2003). Recently, studies have shown evidence of spatial variation in crop micronutrient concentrations at scales relevant for nutrition, not only for Se (Chilimba et al., 2011; Gashu et al., 2021), but also for calcium, iron, zinc (Joy et al., 2015; Gashu et al., 2021; Watts et al., 2019), and iodine (Watts et al., 2015).

Soil and crop Se concentrations have been explicitly linked with human Se status in Malawi (Ligowe et al., 2020). Additionally, the national micronutrient survey shows the risk of Se deficiency in Malawi is high and spatially dependant (Phiri et al., 2019). Plasma Se concentration is highly responsive to variations in dietary Se intake, which in turn, are likely to be influenced by soil type in populations where food systems are dominated by locally grown foods – as demonstrated by plasma Se concentrations and directly measured dietary Se intakes in Malawi (Hurst et al., 2013). Other factors associated with inter-individual variation in plasma Se concentration include proximity to the lake (i.e. a source of freshwater fish), gender, age, smoking, and inflammation (Combs, 2015; Phiri et al., 2019).

Maize provides over 60% of dietary energy in Malawi (Joy et al., 2015). Spatial variation in the Se concentration of maize grain is, therefore, likely to be nutritionally important and a major driver of Se intake, especially for rural populations (Gashu et al., 2021; Gilbert et al., 2019; Joy et al., 2015). Moreover, the Se concentration of maize grain may also serve as a proxy for the Se content of other crop-based components of the Malawian diet given the predominance of locally grown foods in local food systems (Joy et al., 2015).

The spatial variation of dietary Se intake will depend on many underlying factors that are largely unknown, for instance the distances over which food (principally maize given the amounts consumed) moves between the location of production and the location of consumption. These distances, in turn, will depend on various food system factors, including the proportion of food that is sourced from home production, access to markets, seasonality, etc. (Koppmair et al., 2017; Nandi et al., 2021). Other factors which are known to influence the estimates of dietary Se intake are urban/rural residency and household socioeconomic status (Joy, Kumssa et al., 2015). Given the many factors influencing the typical distance between where foods are grown and where they are sourced for consumption, operationalising georeferenced Se food composition data is challenging. Individuals do not source their foods from one single-point location but from an area (e.g. their plot of land, their village or district market), which can be referred to as the household food-shed or catchment area (Koppmair et al., 2017; Nandi et al., 2021). Therefore, the value of maize grain Se concentration used to estimate dietary Se intake should represent the 'average' Se concentration in maize of each individual (or household) food catchment area.

Furthermore, because collecting georeferenced (i.e. point sampled data) composition data is costly, determining the size of the area which represents where the foods, or in this study maize, are sourced and consumed is needed. Such information will inform the sampling design (e.g. number of samples) required to estimate the 'average' Se content in maize for use in dietary assessment.

Hence, the aim was to study the relationship between plasma Se concentration in women (15-49 years old) and maize grain Se concentration at different spatial aggregations. The objective was to identify the level of aggregation that best described the plasma Se concentration to inform the level of aggregation of georeferenced maize Se composition data for future food composition-food consumption matching, which in turn would improve the accuracy of the dietary assessment of localised dietary Se intakes for populations in Malawi. To the best of our knowledge, this is the first study using geostatistical approaches to inform the spatial aggregation of food composition data.

Method

Geospatial models were fitted using Bayesian statistics and the Integrated Nested Laplace Approximation (INLA). The Stochastic Partial Differential Equations (SPDE) approach was used for modelling the spatial process, after confirming that there was spatial autocorrelation in our outcome variable (plasma Se concentration) as shown in the variogram in Suppl. Fig. 1. These geospatial models were implemented using the R-INLA package in RStudio 2023.06.0 Build 421 powered by R version 4.4.1 (Posit team, 2023; R Core Team, 2023) and is described in detail below. The selection of the maize Se concentration spatial aggregation level was based on sub-national administrative units (i.e. district level and EA level) and on food-shed/ food catchment areas (i.e. buffers around displaced EA centroids of 10-60 km radius).

Data: Demographic and Health Survey (DHS)

Data on individual-level plasma Se concentration and participant demographics were extracted from the 2015-16 Malawi Demographic and Health Survey (MDHS) and the Micronutrient Survey (MDHS-MNS) which is a subsample of the 2015-16 MDHS. In brief, the MDHS is a cross-sectional study with a two-stage stratified sampling design. In the first stage, a total of 850 EAs were selected proportional to population size, based on the Malawi Population and Housing Census of 2008. In the second stage, 30 and 33 households were randomly selected for each urban and rural EA, respectively. For the 2015-16 MDHS-MNS, 35 EAs, were selected randomly from the MDHS EA selection (n=850) for each of the three regions in Malawi, and in the second stage for each EA, 20 and 22 households were randomly selected for urban and rural EAs respectively, of which 9 households were randomly selected for eligible women (15-49 years old). In Malawi, the 2015-16 MDHS-MNS data were collected from mid-December 2015 - February 2016. Further details of the sampling design and data collection and analysis for both 2015-16 MDHS and the MNS can be found elsewhere (National Statistical Office (NSO) et al., 2017; National Statistical Office (NSO) [Malawi] and ICF, 2017; Phiri et al., 2019).

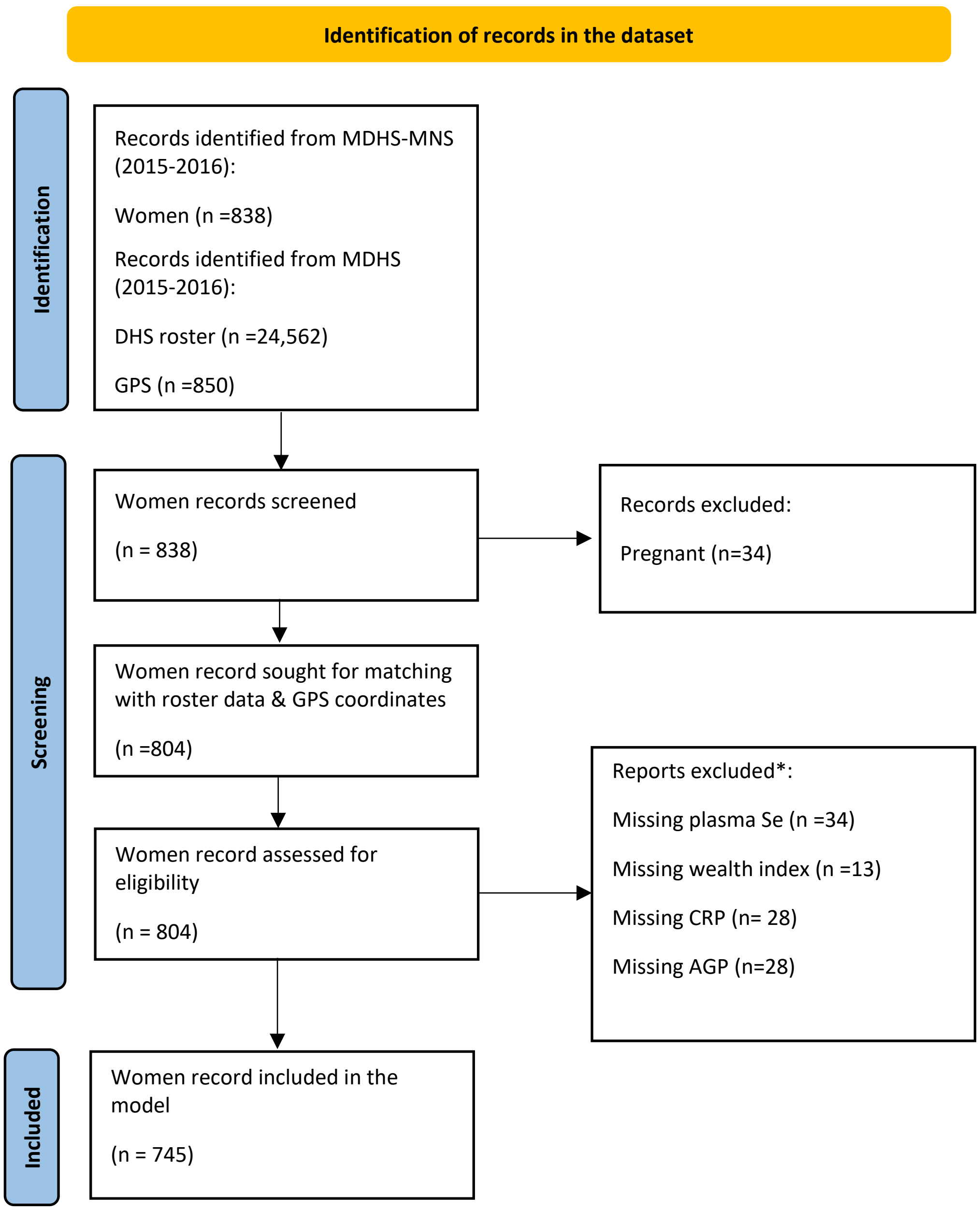


Figure 1: Plasma selenium (Se) concentration in women (15-49 years old) and other covariates extracted from 2015-16 Malawi Demographic and Health Survey (MDHS) and the Micronutrient Survey (MDHS-MNS). Adaptation from PRISMA 2020 flow diagram template for systematic reviews (Page et al., 2021).

Information on blood plasma Se concentration and other individual-level characteristics (age, C-Reactive Protein (CRP), α-1 acid glycoprotein (AGP) and wealth index) were obtained for women aged 15-49 years old from the DHS Program website (Figure 1). Additionally, EA spatial information was provided by DHS which was linked to the household and individual-level data using the EA unique identifier (EA id). The spatial information was provided as the displaced Global Positioning System (GPS) coordinate of the EA centroids. The displacement of the coordinates was 2 km and 5 km for urban and rural EAs, respectively, with an extra 10 km buffer for 1% of the rural EAs. The displacement was always restricted to remain within the true district (administrative unit 2) to provide spatial information while preserving the anonymity of the participants. Thus, women who belong to the same EA would have the same spatial information (i.e. GPS coordinates).

Data: Maize grain selenium concentration

The georeferenced concentration of Se in maize grain in Malawi (sampling points) was generated through the GeoNutrition project which aimed to develop large-scale datasets that allow mapping of factors affecting micronutrient supplies in crops and food systems (Gashu et al., 2021; Kumssa et al., 2022). Paired soil-crop samples (n=1,812) were taken from farmer fields across Malawi, with sampling locations pre-selected to achieve spatial balance. A subset of sample sites (n=820) were co-located with the 2015-16 MDHS (i.e. the EAs selected in the 2015-16 MDHS). In each location, soil and cereal grain samples (maize (n=1,199), sorghum (n=113), rice (n=49) and millet (n=33)), were collected during a single harvest season, i.e. April – June 2018. For each sample, the site GPS location was recorded. Further information on the sampling design, data collection and analysis were previously published (Gashu et al., 2021; Kumssa et al., 2022). The last named publication (Kumssa et al, 2022) provided the georeferenced Se maize data, as the dataset reported in Gashu et al. (2021) excluded 20% of the maize Se samples identified as below limit of detection (<LOD). This data set was complemented by maize grain Se concentration collected in 2009-2010 by Chilimba et al. (2011), which increased the number of samples of maize grain Se concentration, particularly in the Southern region of Malawi in areas where the GeoNutrition survey predominately collected other cereal grains (Suppl. Fig. 2).

In the current study, to predict the maize Se concentration at unsampled locations (i.e. in every square grid of Malawi land) ordinary kriging was used to estimate georeferenced maize grain composition data following the steps outlined in Gashu et al. (2021). In summary, empirical variograms were generated and tested using three estimators (Matheron, 1962), and two robust estimators Cressie and Hawkins (Cressie & Hawkins, 1980), and Dowd (Dowd, 1984), Suppl. Fig. 3b, which reduce the effect of outlying observations on the estimates (Lark, 2000). Variogram models fitted to the estimates were then compared by cross-validation, following the criterion proposed by Lark (2000). On this basis the model fitted to estimates obtained with Dowd's (1984) estimator was selected and used to obtain predictions by ordinary kriging, along with the kriging variance (Suppl. Fig. 4). The analysis and predictions were performed using the *geoR* package (Ribeiro and Diggle, 2001) and additional geostatistical codes in RStudio 2023.06.0 Build 421 powered by R version 4.4.1 (Posit team, 2023; R Core Team, 2023). More information was published in the study protocol, and by Gashu et al. (2021).

Spatial aggregations

Ten different sets of maize Se concentration were calculated for each cluster in the 2015-16 MDHS-MNS as the median of all the predicted maize Se values within the boundaries of the 10 different spatial aggregations. These spatial aggregation consisted in two aggregations based on the administrative boundaries (EA group-level and district-level) which may support decision-making (e.g. in relation to survey planning or agricultural extension), and are consistent with household survey data analysis and reporting (Gething et al., 2015; National Statistical Office, 2017, 2020; National Statistical Office (NSO) et al., 2017), while the others were based on potential cross-boundary food-catchment areas (buffers from 10-60 km radius). All the spatial aggregation levels were larger than the displacement radius of the DHS EAs (2 km and 5 km for urban and rural, respectively) as recommended by DHS documentation and the literature (National Statistical Office (NSO) [Malawi] & ICF, 2017; Perez-Heydrich et al., 2013).

The buffers (10, 15, 20, 25, 30, 40, 50 and 60 km) were generated using the displaced GPS coordinates of the 2015-16 MDHS-MNS EAs as centroids. An EA group aggregation was the set of EAs that a woman likely resided in given the displaced centroid point reported in the dataset. This set was deduced based on the displacement rules imposed by DHS: displacement of the true EA's centroid always falls inside the true district and displacement was up to 2km for urban residency and 5km for rural residency. Figure 2 illustrates these EA group aggregations.

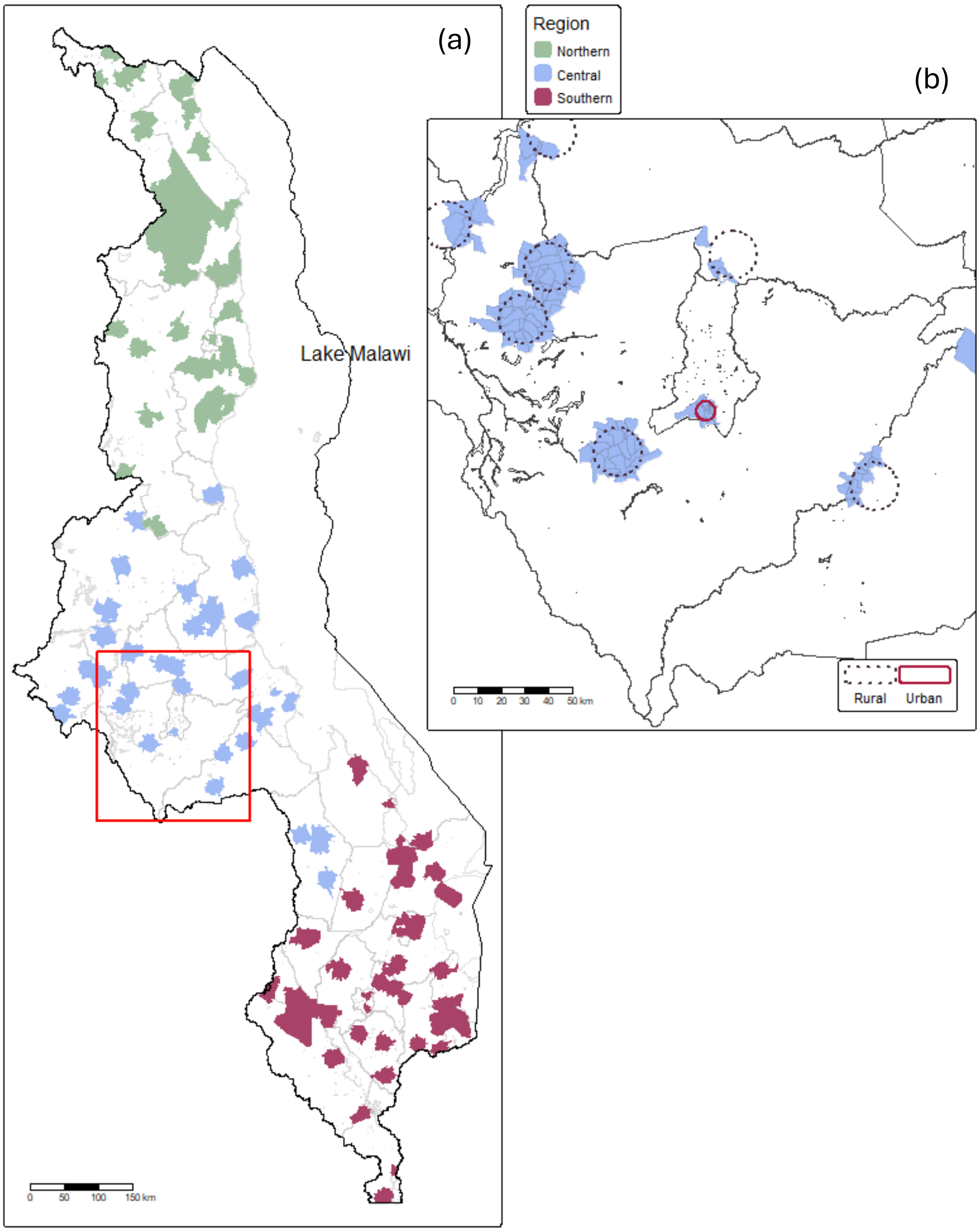


Figure 2: In panel (a) the map of Malawi shows the Enumeration Areas (EAs) included in the EA group aggregation coloured according to their respective region. In panel (b) there is a zoom-in of the Lilongwe district (including the city (urban) and rural sub-districts) showing the EAs included in each EA group based on the displaced coordinates for each EA surveyed in the 2015-2015 Malawi Micronutrient Survey (MDHS-MNS) coloured and sized according to urban (2km buffer) and rural (5km buffer).

District level aggregation was based on the Malawi boundaries obtained from National Statistics Office of Malawi (2018), while boundaries for the smallest administrative units (EAs) were source from MASDAP (2013). These shapefiles were used to generate the spatial aggregations and for visualisation purposes. Finally, the predicted maize grain Se concentration value for each spatial aggregation, calculated as median of all the predicted maize grain Se concentration at each point location within the boundaries of each spatial aggregation, were compared to the observed maize Se values for the same spatial aggregation, with results presented in the Suppl. Fig. 5.

Covariates

Socio-demographic and biological parameters were included as explanatory variables for individual-level plasma Se variation which were obtained from the 2015-16 MDHS. These variables were age, wealth index, residency (urban/rural) and inflammation status (AGP and CRP). The inclusion of these variables was based on previous evidence of their association with plasma Se concentration in Malawi (Galloway et al., 2000; Hurst et al., 2013; Phiri et al., 2019).

In addition to individual-level covariates from the 2015-16 MDHS, distance to the main lakes in Malawi was included. Distance to the lake and/or inland water bodies was shown previously to be a good predictor of fish consumption and plasma Se concentration in Malawi (O'Meara et al., 2021; Phiri et al., 2019; Simmance et al., 2022). Furthermore, fish intake has been identified as one of the main sources of Se in Malawian diets (Joy, Kumssa et al., 2015), and therefore may be a key determinant for plasma Se status in Malawi (Phiri et al., 2019). Distance to main lakes was calculated for each individual based on the shortest Euclidean distance from the (displaced) centroid of their EA to the closest lake border.

Details of each variable used as input in the models, including the source, are provided in Table 1. Further information regarding the individual data sets, cleaning steps and replication of the spatial aggregations can be found in the documentation for reproducing this analysis here.

Table 1. List of the covariates used in the model and their source.

Covariates	**Source**	**Notes**
Maize Se concentration (*mg kg^{-1}*)	Kumssa et al., (2021) Chilimba et al., (2011)	Baseline data (georeferenced point sampled data) for maize selenium aggregations.
Wealth index	2015-16 Malawi Demographic and Health Survey (MDHS) (National Statistical Office & ICF, 2016a)	
Age (*years*)	2015-16 MDHS Micronutrient Survey (MDHS-MNS) (National Statistical Offce & ICF, 2016b)	
C-Reactive Protein (CRP) (*mg L^{-1}*)		
α-1 acid glycoprotein (AGP) (*g L^{-1}*)		

Urbanity	2015-16 MDHS Micronutrient Survey (MDHS-MNS) (National Statistical Office & ICF, 2016b)	
Distance to lake (*Km*)	Calculated from cluster centroids and main lakes in Malawi	As, the shortest Euclidean distance from EA centroids to the closest lake border.

Model framework

The aim of the framework was to understand the spatial relationship between maize grain Se concentration and plasma Se concentration and to identify the spatial aggregation for maize grain Se concentration that best described the individual plasma Se concentration in women (15-49 years old) in Malawi.

A Bayesian framework was implemented using INLA, a computationally very efficient Bayesian solution, which has previously been used to analyse DHS and similar surveys (Gething et al., 2015; Pezzulo et al., 2023; Ruktanonchai et al., 2020). After confirmation of the spatial autocorrelation in the plasma Se concentration (Suppl. Fig. 1), the spatial process was modelled using SPDE and the cluster was included as the independent identical (iid) random effect to account for pseudo-duplication as shown in the Eq. (1).

Equation 1:

$Y_i \sim N\ (\mu_i\ \sigma^2)$, $(i = 1, 2, \ldots, n)$

$y_i = \alpha_c + \beta X_i + \omega_i + \varepsilon_i$

$\omega_i \sim N\ (0, \Sigma_\omega)$

$\varepsilon_i \sim N\ (0, \sigma_e^2)$

Where Y_i is the plasma Se concentration of women (15-49 years old) in Malawi which follow a Gaussian distribution with mean (μ_i) and unknown variance (σ^2), y_i is a vector of i observation of log-transformed plasma Se concentration), random intercept (α_c) for each cluster (c), β are the fixed effects parameters (regression coefficients), X is the matrix that corresponds to the fixed effects (covariates), ω_i is the spatial random effect, and ε_i is the error term. The spatial random effect follows a normal distribution with zero mean and a covariance matrix (Σ_ω). The covariance matrix was model using the Matèrn covariance function. The error term follows a normal distribution with zero mean and unknown variance (σ_e^2).

Ten models were fitted for the different spatial aggregations of the maize Se variable while the other set of the covariates and the random effects (spatial and iid) were kept the same across models. The models were compared using the Deviance Information Criteria (DIC) which is a useful measurement of relative model fit for Bayesian models, such as INLA. Generally, a smaller DIC suggests a better model while a decrease of at least 3 points in DIC provides evidences of model improvement (Blangiardo et al., 2013; Spiegelhalter et al., 2002).

Results

Most of the women (15-49 years old) interviewed in Malawi were living in rural areas (93%) and the median (IQR) concentration of plasma Se concentration of was 76.3 ng mL^{-1} (61.8-94.8), which is below the GPx3 threshold for Se deficiency (<84.9 ng mL^{-1}) as reported in Table 2a. Furthermore, as presented in Table 2b, the median (IQR) value of maize Se concentration for each aggregation was very similar (median range: 0.028-0.031 mg kg^{-1}), with greater variability in aggregation levels with smaller units (i.e. EAs group and buffers from 10-25 km) shown by the larger IQR values and ranges (Figure 3).

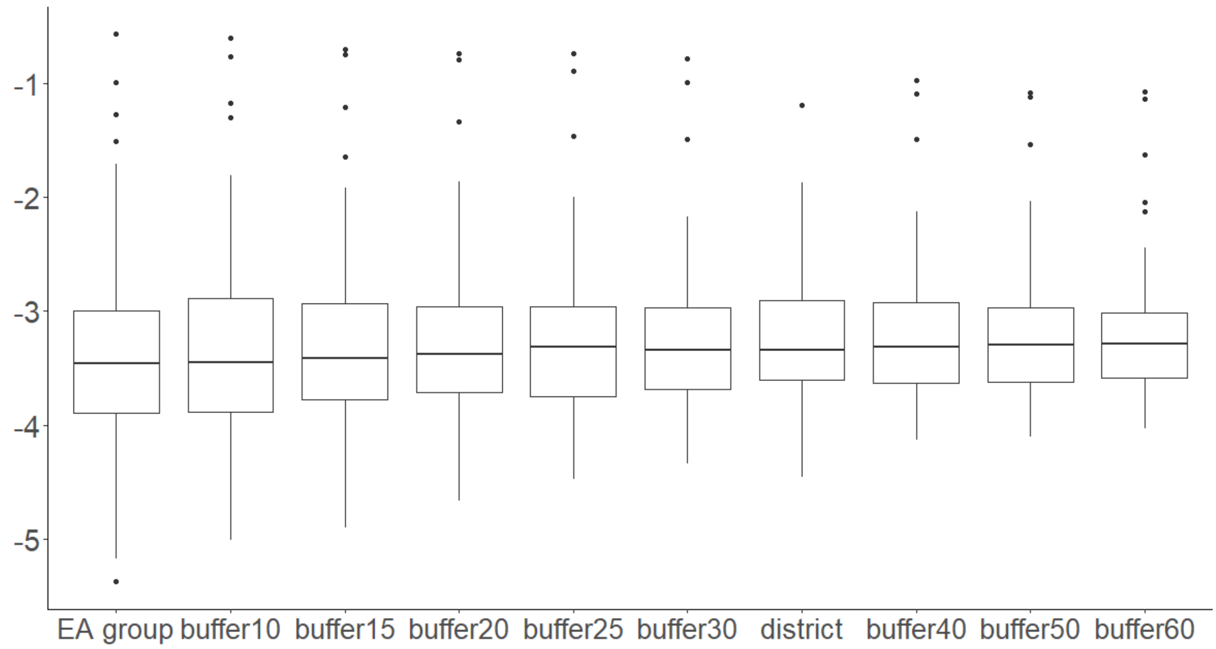


Figure 3. Boxplot of the maize selenium concentration (log-transformed) distribution for each level of aggregation.

The smallest of the aggregation areas was the EA group which also had the highest range of sizes (median (range) area: 213 (13-1,969) km^2) followed by the buffers 10, and 15 km (314 and 706 km^2 respectively) while the largest aggregation levels were 50 and 60 km buffer (Table 2b). Interestingly, the highest range of sizes was district aggregation (median (range) area: 1,225 (14-6,798) km^2) Other summary statistics of the data used in the models can be found in Tables 2a and 2b.

Table 2a. Summary statistics of the data constant across the ten models.

	Units	Median (IQR)
Plasma Se concentration	ng mL^{-1}	76.3 (61.8-94.8)
Age	years	26 (20-36)
CRP	mg L^{-1}	0.71 (0.3-1.82)
AGP	g L^{-1}	0.59 (0.47-0.76)
Distance to lakes	Km	64.84 (3.49-99.16)
Residency		% (N)
Urban		7 (53)
Rural		93 (655)
Wealth index		% (N)
Highest wealth		18 (124)
Higher wealth		26 (185)
Medium wealth		16 (114)
Less wealth		17 (123)
Least wealth		23 (162)

Selenium (Se); C-Reactive Protein (CRP); α-1 acid glycoprotein (AGP)

Table 2b. Summary statistics of the maize selenium concentration and aggregation levels for the 10 models.

	Maize Se conc. Median (IQR)	**Area** Median (range)
Units	mg kg^{-1}	km^2
Buffer10	0.030 (0.021-0.044)	314
Buffer15	0.028 (0.021-0.043)	706
Buffer20	0.030 (0.022-0.044)	1,256
Buffer25	0.031 (0.023-0.042)	1,963
Buffer30	0.030 (0.024-0.042)	2,827
Buffer40	0.031 (0.023-0.040)	5,026
Buffer50	0.031 (0.025-0.038)	7,853
Buffer60	0.031 (0.022-0.037)	11,309
EA Group	0.030 (0.022-0.047)	156 (13-1,969)
District	0.030 (0.023-0.026)	1,225 (14-6,798)

INLA is a Bayesian approach which produces a posterior predictive distribution for each parameter. The significance of the covariates can be evaluated by the mean and the 95% Credible Interval (CI), when the three estimated values, mean, upper and lower bound of the 95% CI are either positive (above zero) or all three negative (below zero). By looking at the mean of the posterior distribution of the coefficients, the results were very similar across all the models, showing a positive association between maize grain Se concentration and plasma Se concentration in women (15-49 years old) in Malawi for all levels of aggregation (Table 3).

Table 3. Summary of probability distribution (mean and 95% Credible Intervals (CI)) of the covariates coefficient for the 10 models tested.

	EA group			District			10km buffer			15km buffer			20km buffer		
	Mean	**Lower 95% CI**	**Upper 95% CI**	**Mean**	**Lower 95% CI**	**Upper 95% CI**	**Mean**	**Lower 95% CI**	**Upper 95% CI**	**Mean**	**Lower 95% CI**	**Upper 95% CI**	**Mean**	**Lower 95% CI**	**Upper 95% CI**
Intercept	4.7017	-20.617	30.0204	4.7156	-20.6047	30.0359	4.6543	-20.6645	29.973	4.7672	-20.5521	30.0864	4.8847	-20.4349	30.2044
Maize Se conc.*	**0.2253**	**0.1518**	**0.2983**	**0.2674**	**0.1377**	**0.3952**	**0.2294**	**0.1516**	**0.3068**	**0.2542**	**0.1628**	**0.3457**	**0.296**	**0.189**	**0.4057**
Wealth index 1	0.9399	-24.3752	26.255	0.9454	-24.3698	26.2605	0.9311	-24.384	26.2462	0.9545	-24.3606	26.2696	0.9794	-24.3357	26.2946
Wealth index 2	0.9252	-24.3899	26.2403	0.9264	-24.3888	26.2416	0.9159	-24.3992	26.231	0.9386	-24.3765	26.2538	0.9623	-24.3529	26.2774
Wealth index 3	0.9458	-24.3693	26.2609	0.9466	-24.3685	26.2618	0.9363	-24.3789	26.2514	0.9581	-24.3571	26.2732	0.981	-24.3341	26.2962
Wealth index 4	0.9065	-24.4086	26.2216	0.9087	-24.4065	26.2239	0.8973	-24.4178	26.2124	0.9191	-24.396	26.2342	0.942	-24.3732	26.2571
Wealth index 5	0.9843	-24.3308	26.2994	0.9885	-24.3267	26.3037	0.9737	-24.3414	26.2888	0.9968	-24.3183	26.312	1.02	-24.2951	26.3352
Age	**0.0025**	**0.0008**	**0.0042**	**0.0026**	**0.0009**	**0.0042**	**0.0025**	**0.0008**	**0.0042**	**0.0025**	**0.0008**	**0.0042**	**0.0025**	**0.0008**	**0.0042**
CRP*	-0.0048	-0.0172	0.0076	-0.0053	-0.0177	0.0071	-0.0049	-0.0173	0.0075	-0.005	-0.0173	0.0074	-0.0051	-0.0175	0.0073
AGP*	0.0144	-0.0307	0.0595	0.0148	-0.0304	0.06	0.0148	-0.0303	0.0599	0.0147	-0.0304	0.0598	0.0145	-0.0307	0.0597
Residence (Rural)	**-0.1479**	**-0.2818**	**-0.0144**	-0.1354	-0.2829	0.0115	-0.1112	-0.2461	0.0231	-0.1058	-0.2432	0.031	-0.0876	-0.2247	0.0487
Distance to lake*	**-0.0442**	**-0.0863**	**0.0000**	-0.0368	-0.0816	0.0112	-0.0424	-0.085	0.0026	**-0.049**	**-0.0919**	**-0.0037**	**-0.0515**	**-0.0947**	**-0.0058**

	25km buffer			30km buffer			40km buffer			50km buffer			60km buffer		
	Mean	**Lower 95% CI**	**Upper 95% CI**	**Mean**	**Lower 95% CI**	**Upper 95% CI**	**Mean**	**Lower 95% CI**	**Upper 95% CI**	**Mean**	**Lower 95% CI**	**Upper 95% CI**	**Mean**	**Lower 95% CI**	**Upper 95% CI**
Intercept	4.8974	-20.4228	30.2176	4.9495	-20.3711	30.2701	4.9364	-20.3852	30.258	4.9804	-20.3425	30.3034	5.0248	-20.2996	30.3492
Maize Se conc.*	**0.3123**	**0.189**	**0.4403**	**0.3398**	**0.2006**	**0.487**	**0.3344**	**0.1731**	**0.5019**	**0.3361**	**0.1565**	**0.5211**	**0.3454**	**0.1456**	**0.5555**
Wealth index 1	0.9827	-24.3325	26.2978	0.9929	-24.3223	26.3081	0.9902	-24.3251	26.3054	0.9986	-24.3166	26.3139	1.0073	-24.308	26.3226
Wealth index 2	0.9642	-24.351	26.2793	0.9746	-24.3405	26.2898	0.9716	-24.3436	26.2869	0.9801	-24.3352	26.2954	0.9891	-24.3263	26.3044
Wealth index 3	0.9834	-24.3317	26.2986	0.994	-24.3212	26.3092	0.9913	-24.3239	26.3066	1.0002	-24.3151	26.3155	1.009	-24.3064	26.3243
Wealth index 4	0.9441	-24.3711	26.2593	0.9543	-24.3609	26.2695	0.9517	-24.3636	26.2669	0.9607	-24.3546	26.2759	0.9696	-24.3457	26.2849
Wealth index 5	1.023	-24.2921	26.3382	1.0336	-24.2816	26.3487	1.0316	-24.2836	26.3468	1.0408	-24.2744	26.3561	1.0499	-24.2655	26.3652
Age	**0.0025**	**0.0009**	**0.0042**	**0.0026**	**0.0009**	**0.0042**	**0.0026**	**0.0009**	**0.0042**	**0.0026**	**0.0009**	**0.0043**	**0.0026**	**0.0009**	**0.0043**
CRP*	-0.0051	-0.0175	0.0073	-0.0051	-0.0175	0.0073	-0.0049	-0.0173	0.0075	-0.0048	-0.0172	0.0076	-0.0048	-0.0172	0.0076
AGP*	0.0142	-0.031	0.0593	0.0142	-0.0309	0.0593	0.0131	-0.0321	0.0583	0.0121	-0.0331	0.0573	0.0125	-0.0328	0.0577
Residence (Rural)	-0.0673	-0.2085	0.0729	-0.0701	-0.2124	0.0712	-0.0801	-0.228	0.0671	-0.0871	-0.2368	0.0619	-0.0913	-0.2421	0.0587
Distance to lake*	**-0.0499**	**-0.0938**	**-0.0033**	**-0.0478**	**-0.092**	**-0.0004**	-0.0477	-0.0927	0.0008	**-0.0514**	**-0.0974**	**-0.002**	**-0.0536**	**-0.1006**	**-0.0026**

*Indicate log transformed variables. CRP: C-Reactive Protein. AGP: α-1 acid glycoprotein.

In addition, the CIs provide information about the certainty of the results. Slightly larger means and wider 95% CIs for maize Se concentration coefficients were found in larger aggregation levels (Figure 4). For instance, the coefficients of the maize grain Se concentration (log-transformed) were 0.225 (95% CI 0.152, 0.298) and 0.267 (95% CI 0.138, 0.395) for EA group (smallest aggregation area) and district (large aggregation area), respectively, indicating a small increase in the uncertainty with an increase in area size. Similar trends can be seen in Figure 4 for 10 km and 60 km buffers, i.e. 0.229 (95% CI 0.152, 0.307) and 0.345 (95% CI 0.146, 0.555), respectively. Two other covariates were also associated with plasma Se concentration: age, which was positively and equally associated with plasma Se concentration in all models, and distance to the lakes. The latter had a very small negative association with plasma Se for all levels of aggregation, except for district level, 10 and 40 km buffer. Additionally, in only the EA group aggregation level model, rural residency was also negatively associated with the plasma Se concentration in women (15-49 years old), i.e. β_{10}= -0.148 (95% CI -0.282, -0.014). The other three covariates (AGP, CRP, wealth index) did not have any effect on the plasma Se estimates (Table 3).

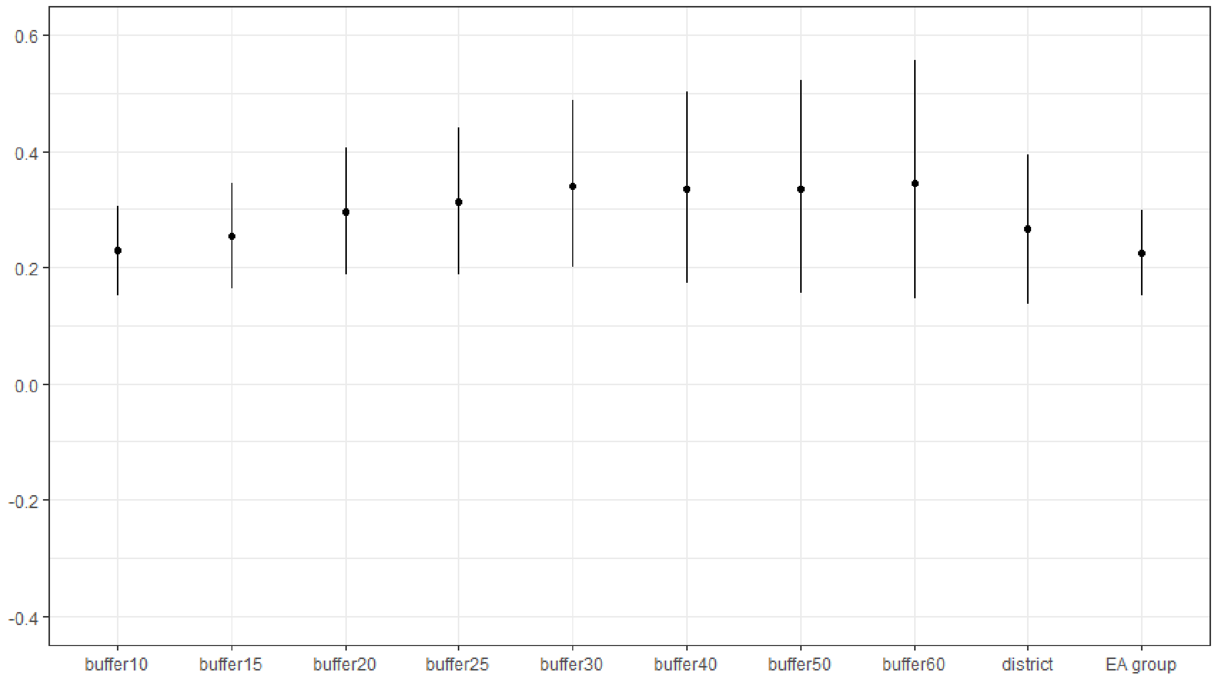


Figure 4. Summary of the posterior distribution (the dot represents the mean and the bars 95% Credible Intervals) of the maize selenium aggregation (log-transformed) coefficient for all the ten models.

All the models regardless of the spatial aggregation levels showed similar model fit according to the DIC, with a non-significant decrease in performance as the aggregation size increased (Figure 5).

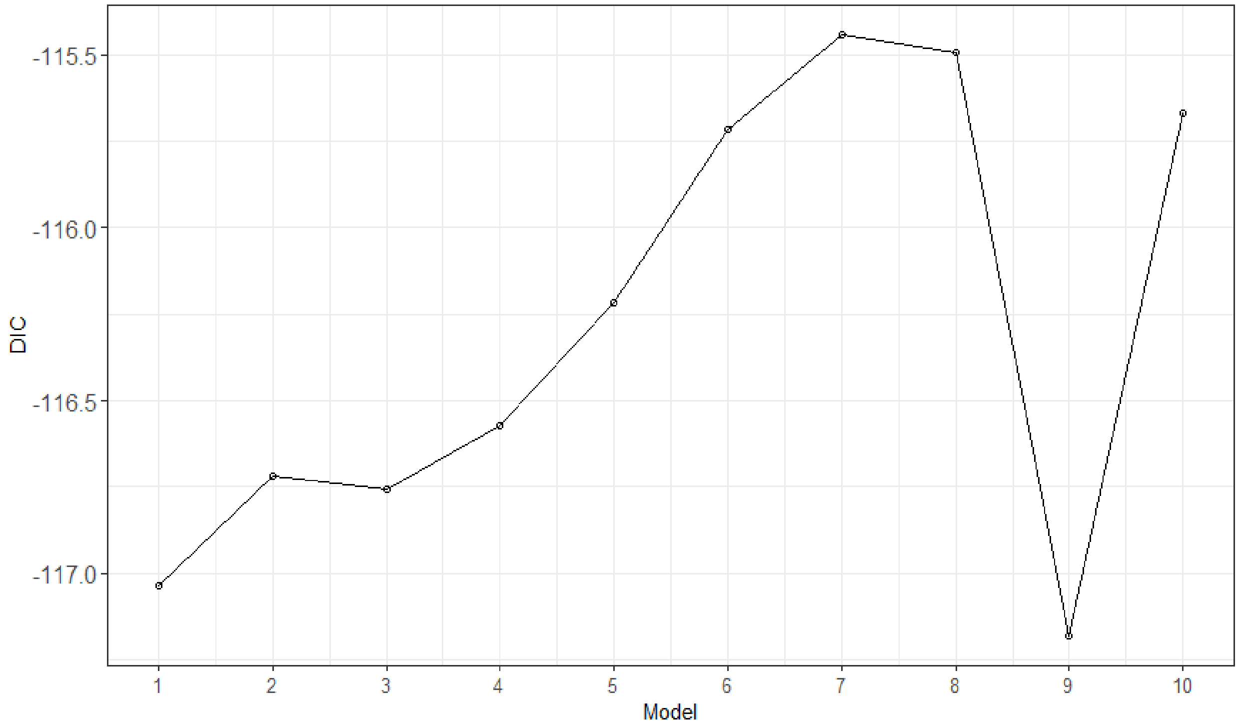


Figure 5. Deviation Information Criterion (DIC) of all ten models. 1=10km buffer, 2=15km buffer, 3=20km buffer, 4=25km buffer, 5=30km buffer, 6=40km buffer, 7=50km buffer, 8=60km buffer, 9=Enumeration Area group level, 10=district level.

Finally, the mean and standard deviation of the residual spatial variation showed marked differences between the models (Suppl. Fig. 6). All the models showed residual spatial effect shown by the different colour shades. However, smaller aggregation levels (e.g. EA group and 10 or 20 km buffer) presented smoother surfaces, and smaller standard deviation, suggesting smaller area aggregations of maize grain composition accounted better for the spatial effects on variation in plasma Se concentration. Moreover, high standard deviations including around the sampled data can be seen for district level and >40 km buffer aggregations which indicate that more granular data/information is needed to accurately predict the true value of plasma Se concentration.

Discussion

Spatial variation in maize grain Se concentration as a key driver of plasma Se concentration

This study evaluated different levels of spatial aggregation for high resolution maize grain Se concentration in Malawi. Subregional maize grain Se concentration aggregations showed a positive association with plasma Se concentration in women (15-49 years old) in Malawi. These results confirmed the importance of using subregional maize grain Se concentration when estimating dietary

Se intake, which are explained by the strong interlinkages between soil type and geochemistry, crops and human Se status, as other authors have reported previously for Malawi and elsewhere (Belay et al., 2020; Fairweather-Tait et al., 2011; Ligowe et al., 2020).

Importantly, by comparing the different spatial aggregations of maize grain Se concentration with plasma Se concentration, the EA group aggregation has the smallest DIC and 95% CI suggesting a higher explanatory potential within this context. Overall, smaller aggregation levels performed better, suggesting the importance of high spatial resolution data for improving the estimates of dietary Se intakes and inadequacy risks in Malawi. Small aggregation levels may reflect better the concentration of Se in maize grain consumed due to the highly localised food systems. For instance, in Malawi >80% of households reported engaging in agriculture of which >90% grow maize (Government of Malawi, 2024; National Statistical Office, 2020). Hence, for most households in Malawi, the majority of maize consumed is likely to be own-produced or purchased from near-by markets which in turn is likely to be grown in similar soil and environmental conditions, yielding proximal Se concentration in the grain. Larger aggregation areas lose detail (i.e. smoothing) as diverse maize Se concentrations are summarised to a single value which may not accurately capture local differences. Given the wide range of Se concentration values across space found by Gashu and colleagues (2019) and the frequency and quantity of maize consumption, small area aggregation are likely to influence the dietary Se estimates when combining food consumption and composition data. Additionally, it may be useful for identifying hotspots or subnational areas with heightened risk of dietary Se inadequate intake.

In this study, from all the socio-demographic and biological parameters included in the models as covariates, age was the only one positively associated with plasma Se concentration in all the models. This contrasts with previous studies showing a negative association with age (Ghayour-Mobarhan et al., 2005; Lloyd et al., 1983). This discrepancy could be due to differences in the age range of participants in these studies. In the previous studies, participants were all >55 years old which is above the age of the women included in this study (Ghayour-Mobarhan et al., 2005; Lloyd et al., 1983).

Distance to the main lakes in Malawi was negatively associated with plasma Se concentration in most of the models (with the exception of district and 10 km buffer). This result is consistent with previous research which showed distance to the lake is a good proxy for fish consumption in Malawi which, in turn, has been reported as one the main source of Se in Malawi (Joy, Kumssa et al., 2015; O'Meara et al., 2021; Phiri et al., 2019; Simmance et al., 2022).

Rural residency was negatively associated with plasma Se concentration only for the EA group aggregation level model. Higher prevalence of Se deficiency among rural than urban women (15-49 years old) in Malawi was previously reported by Phiri and colleagues (2019), which may be due to more diversified food acquisition patterns in urban than rural environments. Joy, Kumssa et al. (2015) reported observed substantial differences in food group consumption between rural and urban areas for Se-rich foods including animal products and fish. The EA group model was the only model that was able to capture these nuances which could be explained by the differences in area size when comparing rural and urban clusters (Suppl. Fig. 7). The area sizes were defined by the GPS coordinates displacement (2 km for urban and 5 km for rural), which led to overall smaller EA group aggregation area sizes in urban than in rural settings. Hence, further research is needed to explore the impact of different aggregation levels in urban and rural areas, and the impact of different food acquisition patterns respectively.

Other parameters, such as wealth index and the inflammation markers (CRP and AGP) were not associated with plasma Se concentration. The lack of association with wealth index could be due to the strong spatial dependency associated with soil and environmental characteristics, hence regardless of the wealth index, people living in high maize Se concentration areas would have higher dietary Se intakes, and in consequence higher plasma Se concentrations than people living in low maize Se concentration areas. The inflammation effect was weak and inconsistent which could be, partially, due to the overall low inflammation status of the adult population (<13%) (National Statistical Office (NSO) et al., 2017). Furthermore, although Se plays a role in inflammation and inflammatory response, the association of AGP and CRP with plasma Se remains unclear (Galloway et al., 2000; Hariharan & Dharmaraj, 2020; Huang et al., 2012).

Comparison of the different levels of aggregation and their meaning for public health nutrition

Overall, the models with smaller maize Se concentration aggregation levels (e.g. EA group and 10 km buffer) explained better the changes in plasma Se concentration than those with large aggregation areas (e.g. district and 60 km buffer), with the EA group level model showing the smallest DIC. Despite that, the difference in DIC was inconclusive, and thus, the selection of the level of aggregation would depend on the objective of the study, the availability of georeferenced maize grain Se concentration data to produce small area aggregations, and the expertise and capacity to perform technically complex aggregations.

The aggregations over small areas may reflect similar soil and environmental factors influencing the maize grain Se concentration but also culturally close communities which may farm, trade and/or exchange food and hence have similar dietary Se intake. Thus, in Malawi, lower intra-cluster variation in plasma Se concentration was found compared to inter-cluster variation (Phiri et al., 2019). This is particularly true for women who are less likely to eat outside their home and/or move outside their communities (Craig et al., 2023).

However, due to the costs and effort required to generate such datasets, the sampling effort used in this study may not be available for other foods and contexts. Hence, in order to identify the upper level of aggregation that could be used to produce FCT data, regional maize grain Se concentration aggregation was tested yielding non-significant results (Suppl. Fig. 8). Therefore, for future data collection and reporting in FCTs, the district level aggregation for maize could be a good option, such as it provides spatial variation information with the least sampling burden.

Further research on other important crops and contexts should be carried out to determine whether similar association between plasma Se concentration and the level of aggregation found for Se concentration in maize grain could be of use in other context with similarly localised food systems. In addition, spatial techniques could be explored for combining very dense sampling datasets, such as the maize Se concentration used in this study, with sparse Se concentration values in a variety of foods to disaggregate their Se concentration values into smaller areas. The use of the concentration of Se in one cereal crop (wheat) to help estimate the Se concentration in another (teff) has been demonstrated previously (Gashu et al., 2020).

The study findings may be valuable for the design of future agronomic biofortification programmes in Malawi. Agronomic biofortification involves adding small quantities of Se to soil-applied fertilisers, leading to enriched Se content in maize. This is a promising approach to tackling Se deficiency in Malawi (Chilimba et al., 2012; Joy et al., 2021) and could be deployed through area-specific fertiliser recommendations – in line with national fertiliser strategies (Ministry of Agriculture, Malawi Government, 2021). Based on the current study findings, small area targeting of Se fertilisers may be an effective and efficient approach to alleviating Se deficiencies, combined with small area monitoring and evaluation.

Strengths and limitations

The strengths of this study are that methods are replicable and adaptable to other staple crops and settings, providing that the data is available. For instance, similar models could be tested for Ethiopia using the high resolution teff Se concentration published by Gashu and colleagues and the serum Se concentration collected by Belay et al. (2020). All the code, decisions and documentation are provided in an open repository in GitHub. In addition, to the best of our knowledge this is the first time that Bayesian spatial statistics are used to identify the level of aggregation of mineral composition data for use in nutritional assessment.

The main limitation is that the crop and plasma data were collected in different time periods, and hence time differences may have affected the results. Moreover, plasma Se data was collected during the 'lean season' where the home-grown maize stock are likely to be at their lowest, and hence household may be turning into increase maize purchases or coping mechanism, such as eating less, relaying on family/social networks, or eating less preferred food. Similarly, the study was focused on women (15-45 years old), who may have different dietary pattern (e.g. inequalities in household food distribution) or may eat more locally than men (Craig et al., 2023).

Another limitation is the proportion of samples with Se concentration <LOD, which in the main maize grain Se concentration dataset was high (~20%). The accuracy of those values is lower than the rest of the dataset and could influence the maize Se concentration predictions. However, the data were retained because the exclusion of those values would have produced a 'left-censored' dataset (i.e. a dataset with all the data points removed below certain value, usually the LOD) which may have further affected the kriging predictions leading to higher (over-estimated) values (Kumssa et al., 2022; Palarea-Albaladejo & Martín-Fernández, 2015). Similarly, the smaller number of maize samples collected in the Southern region led to greater uncertainty in that area as shown by the greater kriging variance (Suppl. Fig. 4b).

As mentioned previously, distance to the lakes was modelled as a proxy for fish consumption, however, the quantity and frequency of fish consumption in Malawi varies among populations living in areas at equivalent distance from the lake (Bartley et al., 2020; O'Meara et al., 2021). Another factor influencing the dietary Se intake is the fish species consumed as Se concentration can vary from <20 µg to >100 µg per 100 g fresh edible portion in the region (Nölle et al., 2020; Simmance et al., 2021). Differences in the quantity, frequency and type of fish consumed could be due a number of factors such as the availability and affordability of fish, among other factors (de Bruyn et al., 2021; Funge-Smith, 2018; Simmance et al., 2021). In a recent study, Bennett and colleagues (2022) mapped the value chains of two key fish species in Malawi: a small fish that is often consumed whole and dried ('Usipa', *Engraulicypris sardella*) and a large fresh fish ('Chambo', *Oreochromis karongae*). They found that rural communities far from the lake tend to consume small, dried fish when compared with urban areas which are more likely to consume fresh and larger fish, particularly in Lilongwe. Hence, the incorporation of more granular data on fish consumption and/or fish value chains could improve future modelling exercises. Although the location of the women (15-49 years old) were displaced, the DHS SAR11 report (Gething et al., 2015) illustrated that the spatial structure of the displaced locations is retained, particularly when the spatial structure is pronounced, as was the case for plasma Se concentration (Suppl. Fig. 1). Furthermore, the goal of this study was to identify the aggregation level which could be operationalised in FCTs and for assessing dietary Se intake, and hence it is unlikely that <10km or EA group areas would be considered, as discussed previously.

Finally, the presence of the modifiable areal unit problem (MAUP), which refers to the potential bias on the results/data arising from the selection of the spatial scale (i.e. different results when data is aggregated using different sizes and configuration), have likely affected our results, as it is inevitable when using aggregated spatial data (Fotheringham & Wong, 1991). Nevertheless, the fact that the direction of our results was similar for all the different levels of aggregation (scales) and configuration (e.g. administrative units and buffers) strengthen the main conclusion of the study: sub-regional maize grain Se concentration is needed to assess Se status in the population of Malawi (Chen et al., 2022). Furthermore, because the goal was not predicting plasma Se concentration but to identify the model that best describes the observed plasma Se concentration in women (15-49 years old) by comparing the model results and relative model fit, these uncertainties and potential sources of bias in variable values are equal among all models, hence the main conclusions would likely remain the same.

Conclusion

Subnational aggregations of maize Se concentration were the main explanatory variable of the changes in plasma Se concentration in women (15-49 years old) in Malawi based on the mean and 95% CI posterior distribution of the maize Se concentration coefficients in all models. The results of our study indicated that smaller levels of aggregation were slightly better at capturing the differences in plasma Se concentration compared to larger spatial aggregation levels, as shown by the smaller 95% CI and the lower DICs. Hence, we recommend that, when possible, the smallest mean/median maize Se concentration aggregation is used for estimating dietary Se intake in Malawi, as we anticipate that would better capture the 'average' Se concentration in maize for that individual or household food catchment area. Moreover, the largest aggregation that should be used, for instance when compiling a local FCT, for maize Se concentration should be district level. This aggregation would still account for variation in plasma Se in Malawi, despite having higher uncertainty. More research is needed to understand the impact of these small aggregation levels on the estimates of dietary Se intake in Malawi. Furthermore, this modelling approach could be expanded to predict plasma Se concentration using high resolution maize Se concentration data, however the use of other covariates should be further explored, as the covariates used here may be insufficient as other food systems and/or behavioural factors play a role in the plasma Se concentration in women (15-49 years old).

References

Ahmad, S., Bailey, E. H., Arshad, M., Ahmed, S., Watts, M. J., & Young, S. D. (2021). Multiple geochemical factors may cause iodine and selenium deficiency in Gilgit-Baltistan, Pakistan. *Environmental Geochemistry and Health*, *43*(11), 4493–4513. https://doi.org/10.1007/s10653-021-00936-9

Bartley, D. M., Kaunda, E., Bennett, A., Funge-Smith, S., & Taylor, W. (2020). *Improving the fish value chain in Malawi: Linking stakeholders to policy makers and donors*. www.glfc.org/pubs/FisheryMgmtDocs/Fmd20-02.pdf

Belay, A., Joy, E. J. M., Chagumaira, C., Zerfu, D., Ander, E. L., Young, S. D., Bailey, E. H., Lark, R. M., Broadley, M. R., & Gashu, D. (2020). Selenium Deficiency Is Widespread and Spatially Dependent in Ethiopia. *Nutrients*, *12*(6), Article 6. https://doi.org/10.3390/nu12061565

Bennett, A., Rice, E., Muhonda, P., Kaunda, E., Katengeza, S., Liverpool-Tasie, L. S. O., Belton, B., Infante, D. M., Ross, J., Virdin, J., & Gondwe, E. (2022). Spatial analysis of aquatic food access can inform nutrition-sensitive policy. *Nature Food*, *3*(12), Article 12. https://doi.org/10.1038/s43016-022-00642-4

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, *7*, 39–55. https://doi.org/10.1016/j.sste.2013.07.003

Chen, X., Ye, X., Widener, M. J., Delmelle, E., Kwan, M.-P., Shannon, J., Racine, E. F., Adams, A., Liang, L., & Jia, P. (2022). A systematic review of the modifiable areal unit problem (MAUP) in community food environmental research. *Urban Informatics*, *1*(1), 22. https://doi.org/10.1007/s44212-022-00021-1

Chilimba, A. D. C., Young, S. D., Black, C. R., Meacham, M. C., Lammel, J., & Broadley, M. R. (2012). Agronomic biofortification of maize with selenium (Se) in Malawi. *Field Crops Research*, *125*, 118–128. https://doi.org/10.1016/j.fcr.2011.08.014

Chilimba, A. D. C., Young, S. D., Black, C. R., Rogerson, K. B., Ander, E. L., Watts, M. J., Lammel, J., & Broadley, M. R. (2011). Maize grain and soil surveys reveal suboptimal dietary selenium intake is widespread in Malawi. *Scientific Reports*, *1*(1), 72. https://doi.org/10.1038/srep00072

Combs, G. F. (2015). Biomarkers of Selenium Status. *Nutrients*, *7*(4), Article 4. https://doi.org/10.3390/nu7042209

Craig, A., Hutton, C., Musa, F. B., & Sheffield, J. (2023). Bonding, bridging and linking social capital combinations for food access; A gendered case study exploring temporal differences in southern Malawi. *Journal of Rural Studies*, *101*, 103039. https://doi.org/10.1016/j.jrurstud.2023.103039

Cressie, N., & Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, *12*(2), 115–125. https://doi.org/10.1007/BF01035243

de Bruyn, J., Wesana, J., Bunting, S. W., Thilsted, S. H., & Cohen, P. J. (2021). Fish Acquisition and Consumption in the African Great Lakes Region through a Food Environment Lens: A Scoping Review. *Nutrients*, *13*(7), 2408. https://doi.org/10.3390/nu13072408

Dowd, P. A. (1984). The Variogram and Kriging: Robust and Resistant Estimators. In *Geostatistics for Natural Resources Characterization* (pp. 91–106). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-3699-7_6

Ene-Obong, H., Schönfeldt, H. C., Campaore, E., Kimani, A., Mwaisaka, R., Vincent, A., Ati, J. E., Kouebou, P., Presser, K., Finglas, P., & Charrondière, U. R. (2019). Importance and use of reliable food composition data generation by nutrition/dietetic professionals towards solving Africa's nutrition problem: Constraints and the role of FAO/INFOODS/AFROFOODS and other stakeholders in future initiatives. *Proceedings of the Nutrition Society*, *78*(4), 496–505. https://doi.org/10.1017/S0029665118002926

Fairweather-Tait, S. J., Bao, Y., Broadley, M. R., Collings, R., Ford, D., Hesketh, J. E., & Hurst, R. (2011). Selenium in Human Health and Disease. *Antioxidants & Redox Signaling*, *14*(7), 1337–1383. https://doi.org/10.1089/ars.2010.3275

Fotheringham, A. S., & Wong, D. W. S. (1991). The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A: Economy and Space*, *23*(7), 1025–1044. https://doi.org/10.1068/a231025

Funge-Smith, S. J. (2018). Review of the state of the world fishery resources: Inland fisheries. *FAO Fisheries and Aquaculture Circular*.

Galloway, S. P., McMillan, D. C., & Sattar, N. (2000). Effect of the inflammatory response on trace element and vitamin status. *Annals of Clinical Biochemistry*, *37*(3), 289–297. https://doi.org/10.1258/0004563001899429

Gashu, D., Lark, R. M., Milne, A. E., Amede, T., Bailey, E. H., Chagumaira, C., Dunham, S. J., Gameda, S., Kumssa, D. B., Mossa, A. W., Walsh, M. G., Wilson, L., Young, S. D., Ander, E. L., Broadley, M. R., Joy, E. J. M., & McGrath, S. P. (2020). Spatial prediction of the concentration of selenium (Se) in grain across part of Amhara Region, Ethiopia. *The Science of the Total Environment*, *733*, 139231. https://doi.org/10.1016/j.scitotenv.2020.139231

Gashu, D., Marquis, G. S., Bougma, K., & Stoecker, B. J. (2019). Spatial Variation of Human Selenium in Ethiopia. *Biological Trace Element Research*, *189*(2), 354–360. https://doi.org/10.1007/s12011-018-1489-5

Gashu, D., Nalivata, P. C., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A. E., Mossa, A. W., Munthali, M., … Broadley, M. R. (2021). The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*, *594*(7861), Article 7861. https://doi.org/10.1038/s41586-021-03559-3

Gething, P., Tatem, A., Bird, T., & Burgert-Brucker, C. R. (2015). *Creating spatial interpolation surfaces with dhs data* (DHS Spatial Analysis Reports No. 11). ICF International.

Ghayour-Mobarhan, M., Taylor, A., New, S. A., Lamb, D. J., & Ferns, G. a. A. (2005). Determinants of serum copper, zinc and selenium in healthy subjects. *Annals of Clinical Biochemistry*, *42*(Pt 5), 364–375. https://doi.org/10.1258/0004563054889990

Gilbert, R., Benson, T., & Ecker, O. (2019). *Are Malawian Diets Changing? An assessment of nutrient consumption and dietary patterns using household-level evidence from 2010/11 and 2016/17*. Intl Food Policy Res Inst.

Government of Malawi. (2024). *National Agriculture Policy 2024*. https://agriculture.gov.mw/files/medias/national-agriculture-policy-2024-signed-1739192665.pdf

Grande, F., Giuntini, E. B., Lajolo, F. M., & de Menezes, E. W. (2016). How do calculation method and food data source affect estimates of vitamin A content in foods and dietary intake? *Journal of Food Composition and Analysis*, *46*, 60–69. https://doi.org/10.1016/j.jfca.2015.11.006

Greenfield, H., & Southgate, D. A. T. (2003). *Food composition data: Production, management, and use*. FAO.

Hariharan, S., & Dharmaraj, S. (2020). Selenium and selenoproteins: It's role in regulation of inflammation. *Inflammopharmacology*, *28*(3), 667–695. https://doi.org/10.1007/s10787-020-00690-x

Huang, Z., Rose, A. H., & Hoffmann, P. R. (2012). The Role of Selenium in Inflammation and Immunity: From Molecular Mechanisms to Therapeutic Opportunities. *Antioxidants & Redox Signaling*, *16*(7), 705–743. https://doi.org/10.1089/ars.2011.4145

Hurst, R., Siyame, E. W. P., Young, S. D., Chilimba, A. D. C., Joy, E. J. M., Black, C. R., Ander, E. L., Watts, M. J., Chilima, B., Gondwe, J., Kang'ombe, D., Stein, A. J., Fairweather-Tait, S. J., Gibson, R. S., Kalimbira, A. A., & Broadley, M. R. (2013). Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi. *Scientific Reports*, *3*, 1425–1425. Africa-Wide Information. https://doi.org/10.1038/srep01425

Joy, E. J. M., Broadley, M.R., Young, S.D., Black, C.R., Chilimba, A.D.C., Ander, E.L., Barlow, T.S., Watts, M.J. (2015). Soil type influences crop mineral composition in Malawi. Science of the Total Environment 505, 587–595. https://doi.org/10.1016/j.scitotenv.2014.10.038

Joy, E. J. M., Kalimbira, A. A., Sturgess, J., Banda, L., Chiutsi-Phiri, G., Manase, H., Gondwe, J., Ferguson, E. L., Kalumikiza, Z., Bailey, E. H., Young, S. D., Matandika, L., Mfutso-Bengo, J., Millar, K., Niksic, M., Segovia de la Revilla, L., Likoswe, B. H., Phuka, J. C., Phiri, F. P., … Allen, E. (2021). Biofortified Maize Improves Selenium Status of Women and Children in a Rural Community in Malawi: Results of the Addressing Hidden Hunger With Agronomy Randomized Controlled Trial. *Frontiers in Nutrition*, *8*, 788096. https://doi.org/10.3389/fnut.2021.788096

Joy, E. J. M., Kumssa, D.B., Broadley, M.R., Watts, M.J., Young, S.D., Chilimba, A.D.C., Ander, E.L. (2015). Dietary mineral supplies in Malawi: spatial and socioeconomic assessment. BMC Nutrition 1, 42. https://doi.org/10.1186/s40795-015-0036-4

Koppmair, S., Kassie, M., & Qaim, M. (2017). Farm production, market access and dietary diversity in Malawi. *Public Health Nutrition*, *20*(2), 325–335. https://doi.org/10.1017/S1368980016002135

Kumssa, D. B., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Chimungu, J. G., Davis, K., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A., Muleya, P., Munthali, M., … Nalivata, P. C. (2022). Cereal grain mineral micronutrient and soil chemistry data from GeoNutrition surveys in ethiopia and malawi. *Scientific Data*, *9*(1), Article 1. https://doi.org/10.1038/s41597-022-01500-5

Lark, R. M. (2000). A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science*, *51*(1), 137–157. https://doi.org/10.1046/j.1365-2389.2000.00280.x

Ligowe, I. S., Phiri, F. P., Ander, E. L., Bailey, E. H., Chilimba, A. D. C., Gashu, D., Joy, E. J. M., Lark, R. M., Kabambe, V., Kalimbira, A. A., Kumssa, D. B., Nalivata, P. C., Young, S. D., & Broadley, M. R. (2020). Selenium deficiency risks in sub-Saharan African food systems and their geospatial linkages. *Proceedings of the Nutrition Society*, *79*(4), 457–467. https://doi.org/10.1017/S0029665120006904

Lloyd, B., Lloyd, R. S., & Clayton, B. E. (1983). Effect of smoking, alcohol, and other factors on the selenium status of a healthy population. *Journal of Epidemiology and Community Health*, *37*(3), 213–217.

Long, Z., Yuan, L., Hou, Y., Bañuelos, G. S., Liu, Y., Pan, L., Liu, X., & Yin, X. (2018). Spatial variations in soil selenium and residential dietary selenium intake in a selenium-rich county, Shitai, Anhui, China. *Journal of Trace Elements in Medicine and Biology*, *50*, 111–116. https://doi.org/10.1016/j.jtemb.2018.06.019

Matheron, G. (1962). *Traité de géostatistique appliquée*. Éditions Technip.

MASDAP (Malawi Spatial Data Platform). (2013). Population over Enumeration Areas NSO. https://www.masdap.mw/catalogue/#/dataset/126

Ministry of Agriculture, Malawi Government. (2021). *National Fertiliser Policy*. https://npc.mw/wp-content/uploads/2023/10/2021.03-National-Fertiliser-Policy-signed.pdf

Nandi, R., Nedumaran, S., & Ravula, P. (2021). The interplay between food market access and farm household dietary diversity in low and middle income countries: A systematic review of literature. *Global Food Security*, *28*, 100484. https://doi.org/10.1016/j.gfs.2020.100484

National Statistical Office (NSO). (2017). *Fourth Integrated Household Survey (IHS4) 2016-2017. Basic Information Document*.

National Statistics Office of Malawi (2018). Malawi - Subnational Administrative Boundaries. https://data.humdata.org/dataset/20eb8e5b-134d-41d8-a56f-4f358f7faf16/resource/9d0efd98-5cf3-400a-a303-58cc92b1c5e7/

National Statistical Office. (2020). *The Integrated Household Survey (IHS5) 2020, Report*. https://microdata.worldbank.org/index.php/catalog/3818/download/49046

National Statistical Office (NSO), Community Health Sciences Unit (CHSU) [Malawi], Centers for Disease Control and Prevention (CDC), & Emory University. (2017). *Malawi Micronutrient Survey 2015-16*. National Statistical Office. https://dhsprogram.com/pubs/pdf/FR319/FR319.m.final.pdf

National Statistical Office (NSO) [Malawi], & ICF. (2017). *Malawi Demographic and Health Survey 2015-16*. https://dhsprogram.com/pubs/pdf/FR319/FR319.pdf

Neufeld, L. M., Baker, S., Garrett, G. S., & Haddad, L. (2017). Coverage and Utilization in Food Fortification Programs: Critical and Neglected Areas of Evaluation. *The Journal of Nutrition*, *147*(5), 1015S-1019S. https://doi.org/10.3945/jn.116.246157

Nölle, N., Genschick, S., Schwadorf, K., Hrenn, H., Brandner, S., & Biesalski, H. K. (2020). Fish as a source of (micro)nutrients to combat hidden hunger in Zambia. *Food Security*, *12*(6), 1385–1406. https://doi.org/10.1007/s12571-020-01060-9

O'Meara, L., Cohen, P. J., Simmance, F., Marinda, P., Nagoli, J., Teoh, S. J., Funge-Smith, S., Mills, D. J., Thilsted, S. H., & Byrd, K. A. (2021). Inland fisheries critical for the diet quality of young children in sub-Saharan Africa. *Global Food Security*, *28*, 100483. https://doi.org/10.1016/j.gfs.2020.100483

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. BMJ, 372, n160. https://doi.org/10.1136/bmj.n160

Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2015). zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems*, *143*, 85–96. https://doi.org/10.1016/j.chemolab.2015.02.019

Perez-Heydrich, C., Warren, J. L., Burgert, C. R., & Emch, M. (2013). *Guidelines on the use of DHS GPS data*. ICF International. https://dhsprogram.com/pubs/pdf/SAR8/SAR8.pdf

Pezzulo, C., Tejedor-Garavito, N., Chan, H. M. T., Dreoni, I., Kerr, D., Ghosh, S., Bonnie, A., Bondarenko, M., Salasibew, M., & Tatem, A. J. (2023). A subnational reproductive, maternal, newborn, child, and adolescent health and development atlas of India. *Scientific Data*, *10*(1), 86. https://doi.org/10.1038/s41597-023-01961-2

Phiri, F. P., Ander, E. L., Bailey, E. H., Chilima, B., Chilimba, A. D. C., Gondwe, J., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Phuka, J. C., Salter, A., Suchdev, P. S., Watts, M. J., Young, S. D., & Broadley, M. R. (2019). The risk of selenium deficiency in Malawi is large and varies over multiple spatial scales. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-43013-z

Popkin, B. M., Corvalan, C., & Grummer-Strawn, L. M. (2020). Dynamics of the double burden of malnutrition and the changing nutrition reality. *The Lancet*, *395*(10217), 65–74. https://doi.org/10.1016/S0140-6736(19)32497-3

Posit team. (2023). *RStudio: Integrated Development for R.* [Computer software]. Posit Software, PBC. http://www.posit.co/

R Core Team. (2023). *R: A language and environment for statistical computing.* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ribeiro Jr, P., & Diggle, P. (2001). GeoR: A Package for Geostatistical Analysis. *R-NEWS*, *1*, 14–18.

Ruktanonchai, C. W., Nieves, J. J., Ruktanonchai, N. W., Nilsen, K., Steele, J. E., Matthews, Z., & Tatem, A. J. (2020). Estimating uncertainty in geospatial modelling at multiple spatial resolutions: The pattern of delivery via caesarean section in Tanzania. *BMJ Global Health*, *4*(Suppl 5), e002092. https://doi.org/10.1136/bmjgh-2019-002092

Segovia de la Revilla, L., Ferguson, E. L., Dooley, C., Osman, G., Ander, E. L., & Joy, E. J. M. (2023). The availability and geographic location of open-source food composition data used to estimate micronutrient intakes in sub-Saharan Africa: A scoping review. *Journal of Food Composition and Analysis*, *120*, 105322. https://doi.org/10.1016/j.jfca.2023.105322

Simmance, F. A., Kanyumba, L., Cohen, P., Njaya, F., Nankwenya, B., Gondwe, E., Manyungwa, C., Chimatiro, S., Byrd, K., O'Meara, L., Robinson, J., Ahern, M., Wesana, J., Kolding, J., Cowx, I., Gilbert, R., Bunting, S., Nico, G., Franz, N., … Siamudaala, V. (2021). *Sustaining and improving the contribution small-scale fisheries make to healthy and sustainable food systems in Malawi*.

Simmance, F. A., Nico, G., Funge-Smith, S., Basurto, X., Franz, N., Teoh, S. J., Byrd, K. A., Kolding, J., Ahern, M., Cohen, P. J., Nankwenya, B., Gondwe, E., Virdin, J., Chimatiro, S., Nagoli, J., Kaunda, E., Thilsted, S. H., & Mills, D. J. (2022). Proximity to small-scale inland and coastal fisheries is associated with improved income and food security. *Communications Earth & Environment*, *3*(1), Article 1. https://doi.org/10.1038/s43247-022-00496-5

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353

Watts, M. J., Joy, E. J. M., Young, S. D., Broadley, M. R., Chilimba, A. D. C., Gibson, R. S., Siyame, E. W. P., Kalimbira, A. A., Chilima, B., & Ander, E. L. (2015). Iodine source apportionment in the Malawian diet. *Scientific Reports*, *5*(1), 15251. https://doi.org/10.1038/srep15251

Watts, M. J., Middleton, D. R. S., Marriott, A. L., Humphrey, O. S., Hamilton, E. M., Gardner, A., Smith, M., McCormack, V. A., Menya, D., Munishi, M. O., Mmbaga, B. T., & Osano, O. (2019). Source apportionment of micronutrients in the diets of Kilimanjaro, Tanzania and Counties of Western Kenya. *Scientific Reports*, *9*(1), 14447–14447. https://doi.org/10.1038/s41598-019-51075-2

Supplementary Materials

Supplementary Figure 1. Empirical variogram of the plasma selenium concentration in women (15-49 years old) in Malawi.

Supplementary Figure 2. Map of Malawi showing the sampling location points adapted from GeoNutrition project (Gashu et al., 2021; Kumssa et al., 2022) and Chilimba et al, (2011).

Supplementary Figure 2b. Histograms representing the concentration of selenium in different crops collected in Malawi by the GeoNutrition project (Gashu et al., 2021; Kumssa et al., 2022).

Supplementary Figure 3. Observed selenium concentration in maize grain in Malawi (a) and the empirical variograms using the three estimators (b). The colour scale represents the data dived by quartiles of selenium concentration (high = red, mid-high = yellow, mid-low = green, and low =blue).

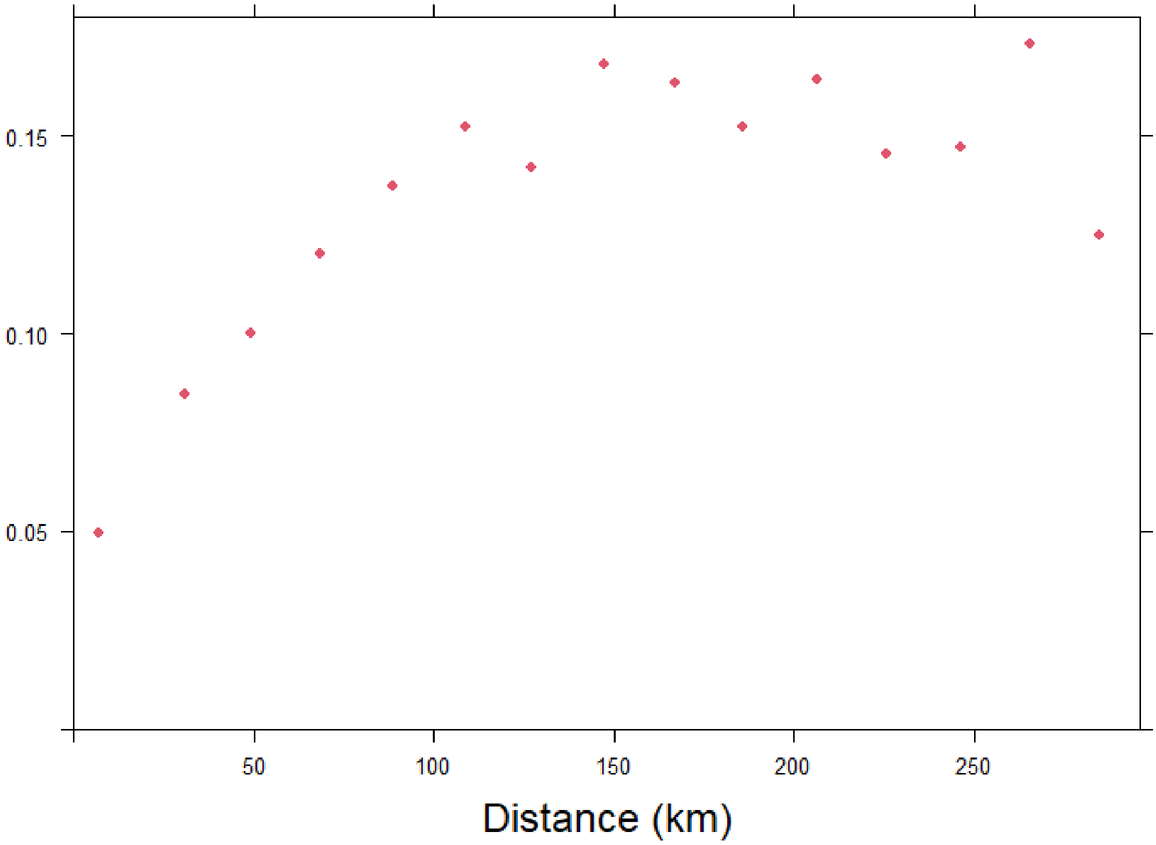
Supplementary Figure 4. Predicted selenium concentration in maize grain in Malawi (a) and its kriging variance (b). The colour scale represents higher concentration (quartile 85, red). The colour scale represents the data dived by quartiles of selenium concentration (high = red, mid-high = yellow, mid-low = green, and low =blue).

Supplementary Figure 5. Each panel shows a plot with each EA mean predicted maize Se concentration and the mean observed maize Se in a) 10 km, b) 15 km, c) 20 km, d) 25 km, e) 30 km, f) 40 km, g) 50 km, h) 60 km, i) EA group and j) district aggregations. All values are log-transformed.

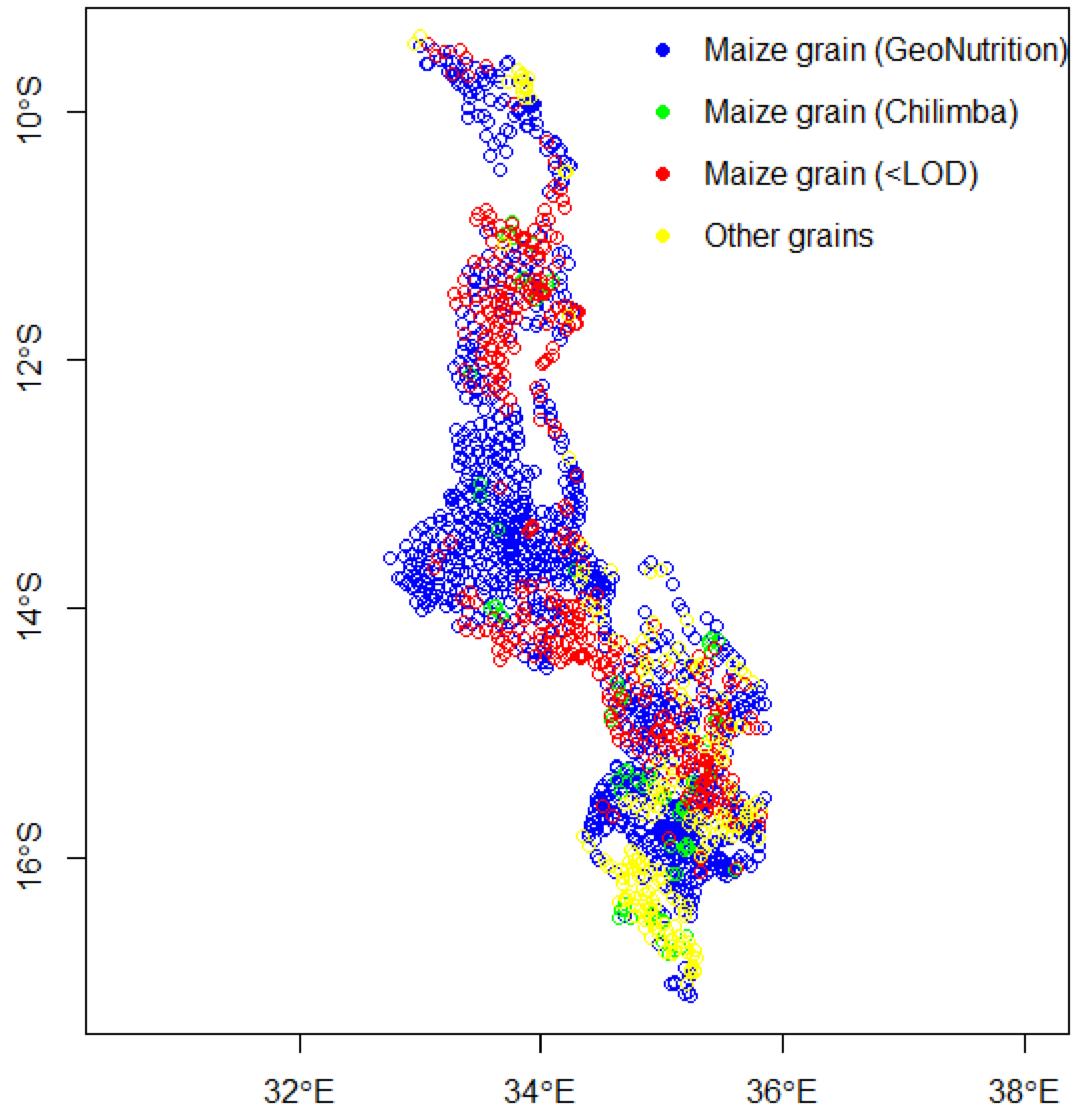
Supplementary Figure 6. Each map of Malawi shows the mean and standard deviation of the residual spatial variation for the a) 10 km, b) 15 km, c) 20 km, d) 25 km, e) 30 km, f) 40 km, g) 50 km, h) 60 km, i) district and j) EA group models.

Supplementary Figure 7. Boxplots representing the area size (in km^2) of the EA group aggregation by rural and urban 2014-2015 MDHS cluster.

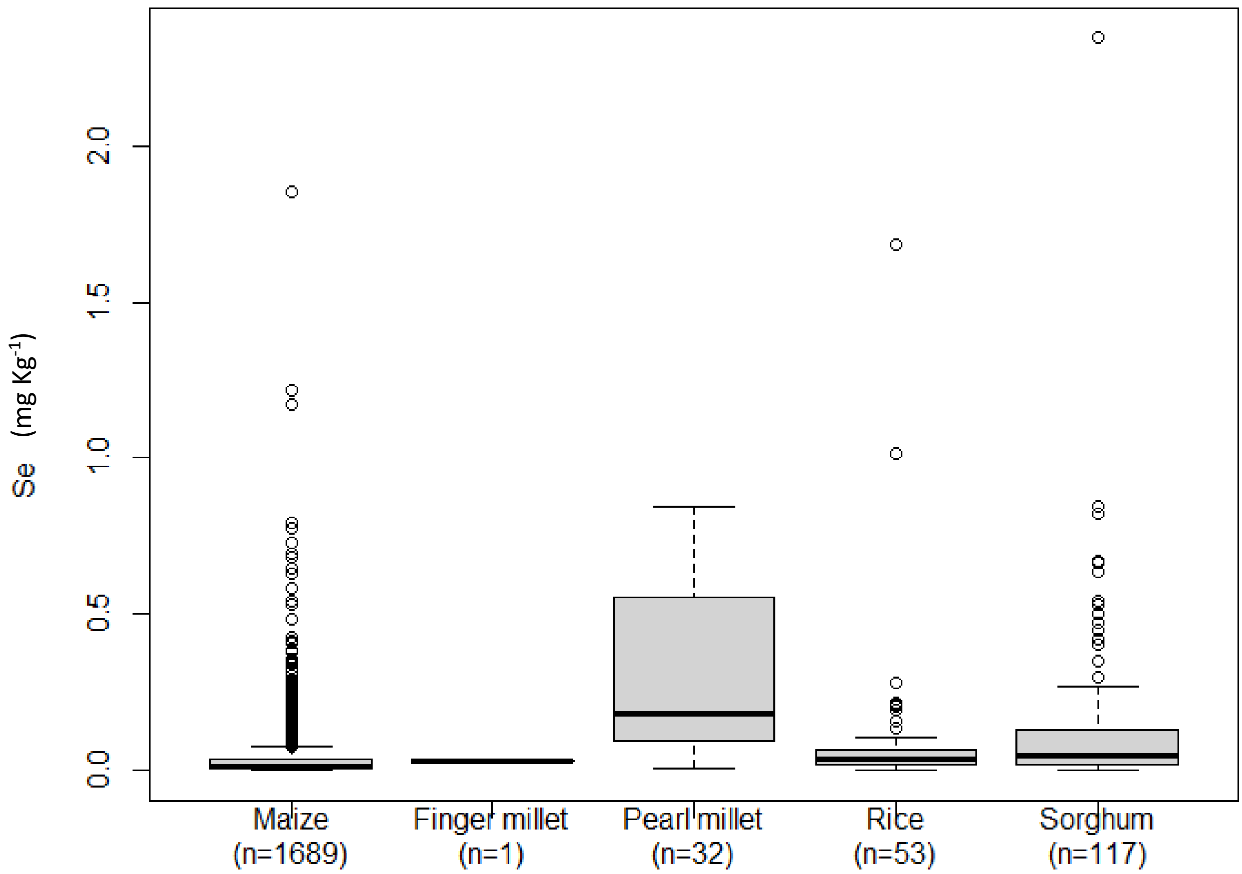
Supplementary Figure 8. Summary of the posterior distribution (the dot represents the mean and the bars 95% Credible Intervals) of the maize selenium aggregation (log-transformed) coefficient for all the ten models and the national-level model.



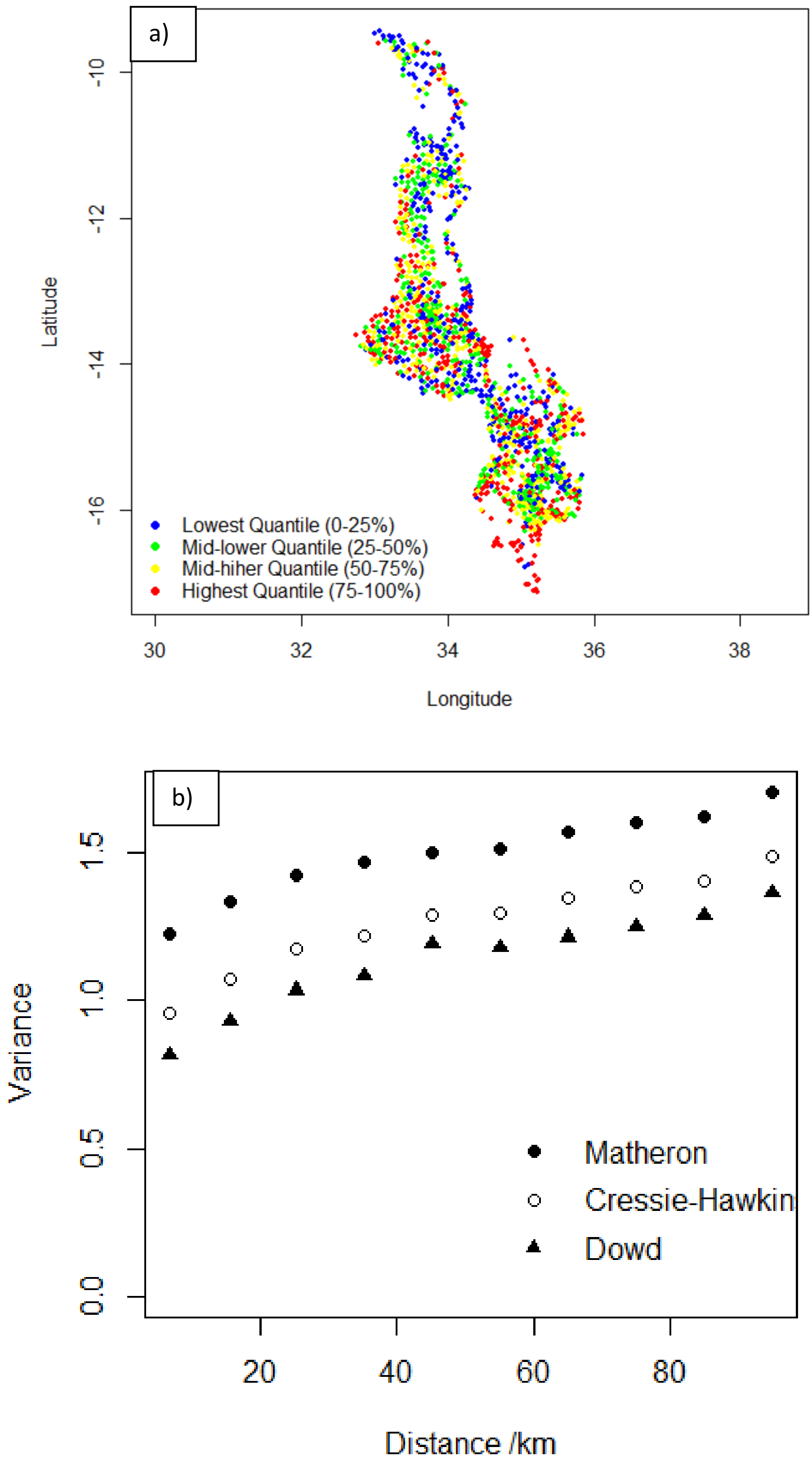
Supplementary Figure 1. Empirical variogram of the plasma selenium concentration in women (15-49 years old) in Malawi.



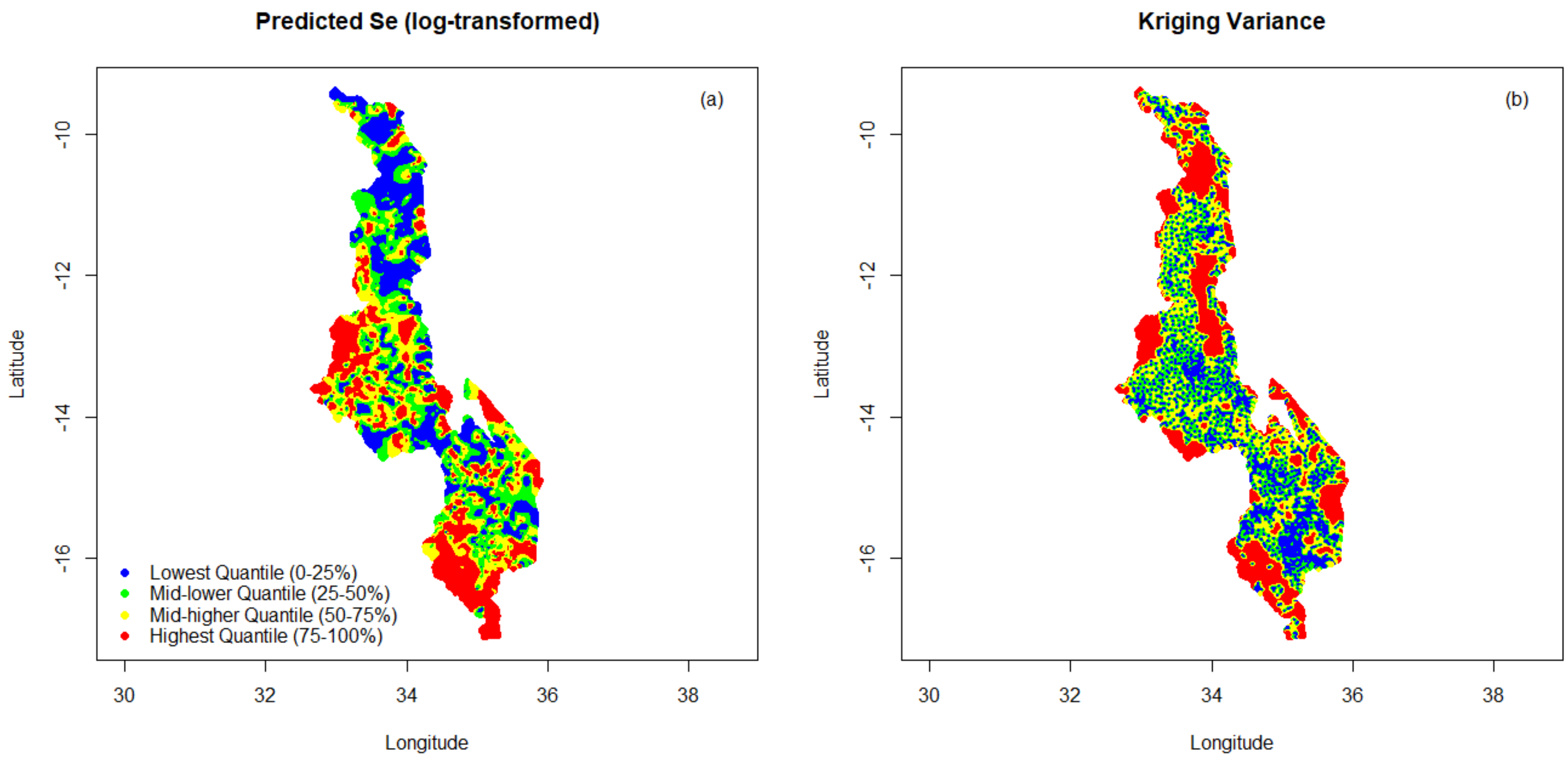
Supplementary Figure 2. Map of Malawi showing the sampling location points adapted from the GeoNutrition project (Gashu et al., 2021; Kumssa et al., 2022) and Chilimba et al, (2011). The colours represent the observed maize selenium concentration by the GeoNutrition project (blue, n= 1,199), and from Chilimba et al. (2011, green, n=85). In red, are the maize selenium concentration below limit of detection (<LOD) (n=404, representing 25% of the sample), and in yellow the "other grains" selenium concentrations collected by the GeoNutrition project (n= 204, representing 11% of the dataset).



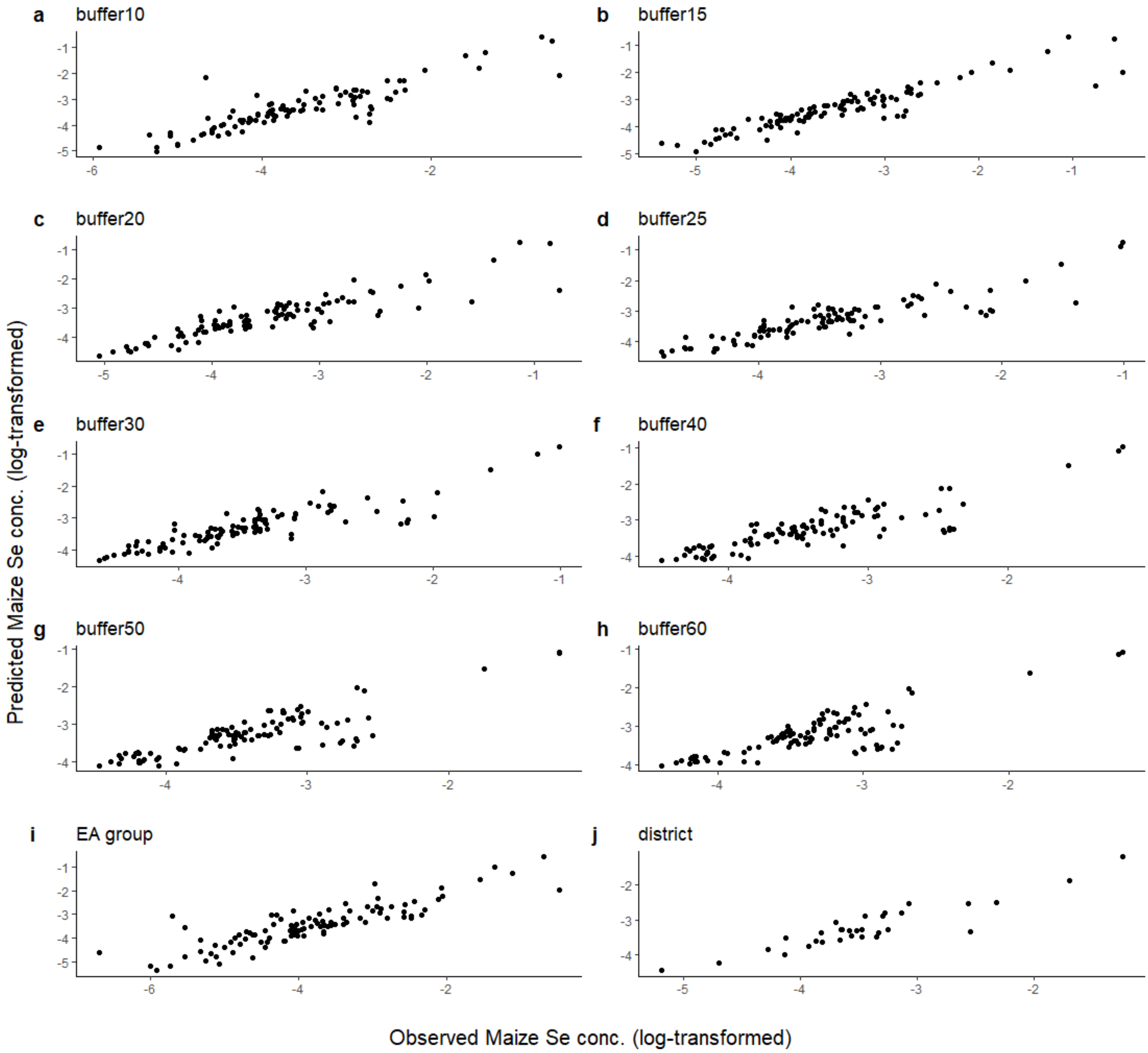
Supplementary Figure 2b. Histograms representing the concentration of selenium (mg Kg^{-1}) in different crops collected in Malawi by the GeoNutrition project (Gashu et al., 2021; Kumssa et al., 2022). In parenthesis are the number of samples for each crop.



Supplementary Figure 3. Observed selenium concentration in maize grain in Malawi (a) and the empirical variograms using the three estimators (b). The colour scale represents the data dived by quartiles of selenium concentration (high = red, mid-high = yellow, mid-low = green, and low =blue).

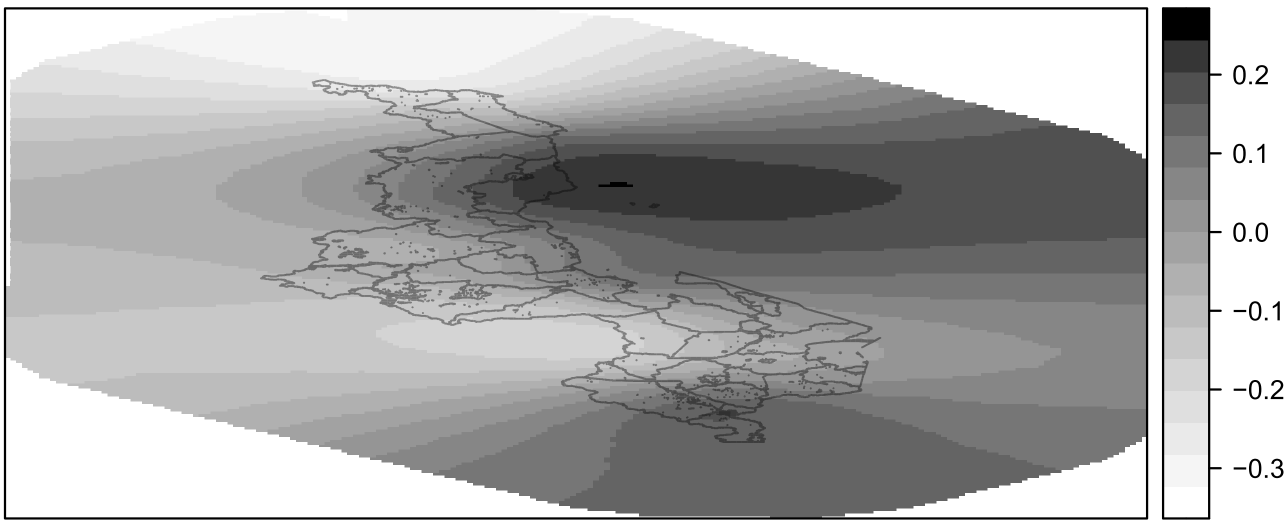


Supplementary Figure 4. Predicted selenium concentration (log-transformed) in maize grain in Malawi (a) and its kriging variance (b). The colour scale represents higher concentration (quartile 85, red). The colour scale represents the data dived by quartiles of selenium concentration (high = red, mid-high = yellow, mid-low = green, and low =blue)

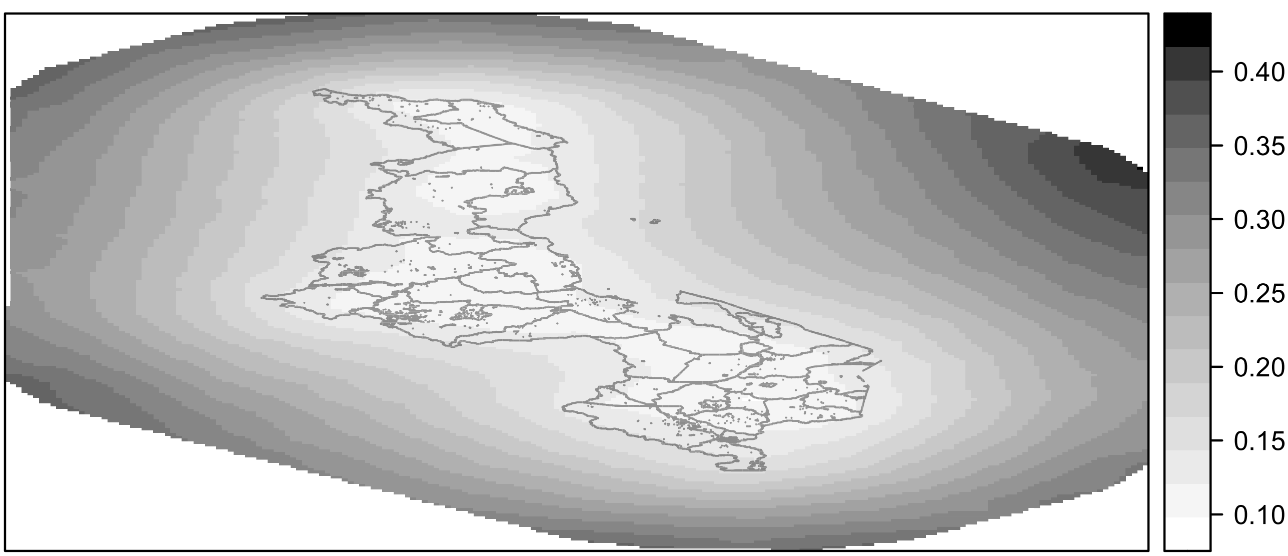


Supplementary Figure 5. Each panel shows a plot with each EA mean predicted maize Se concentration and the mean observed maize Se in a) 10 km, b) 15 km, c) 20 km, d) 25 km, e) 30 km, f) 40 km, g) 50 km, h) 60 km, i) EA group and j) district aggregations. All values are log-transformed.

buffer10

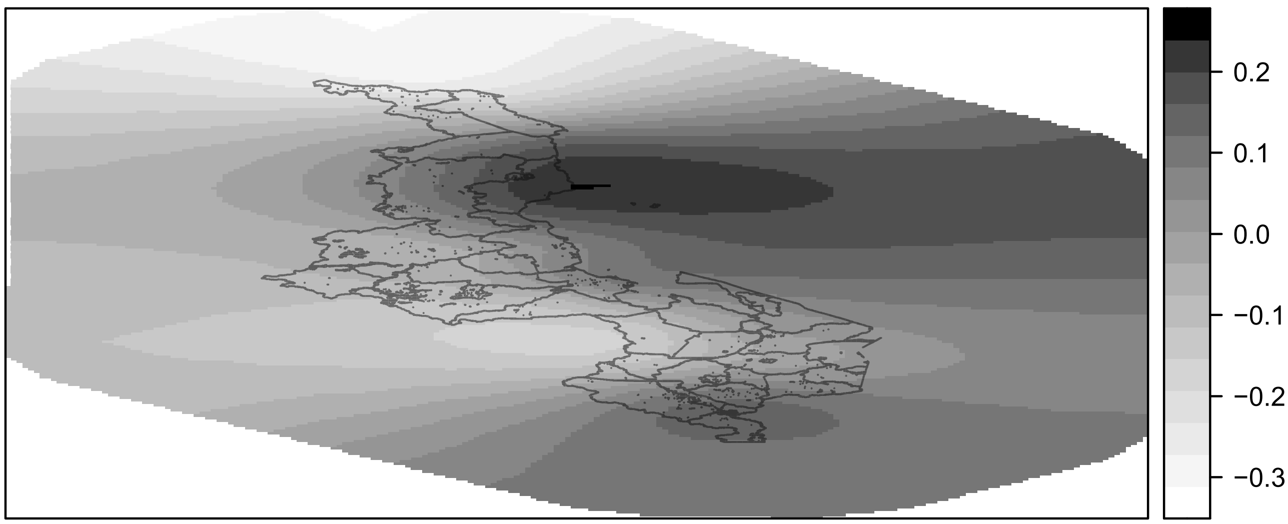


Mean

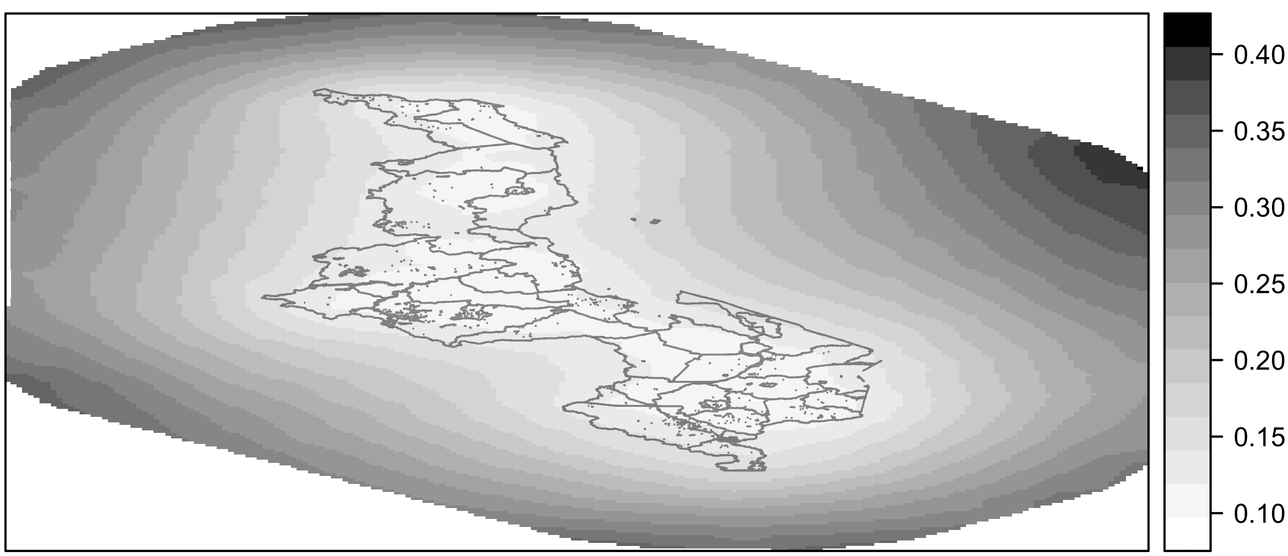


SD

buffer15

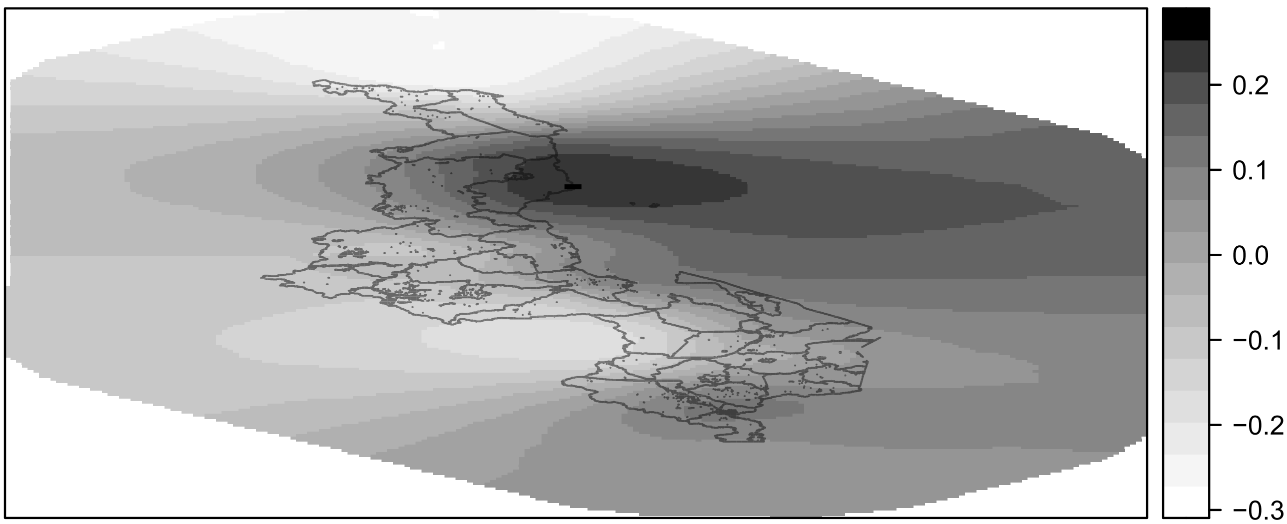


Mean

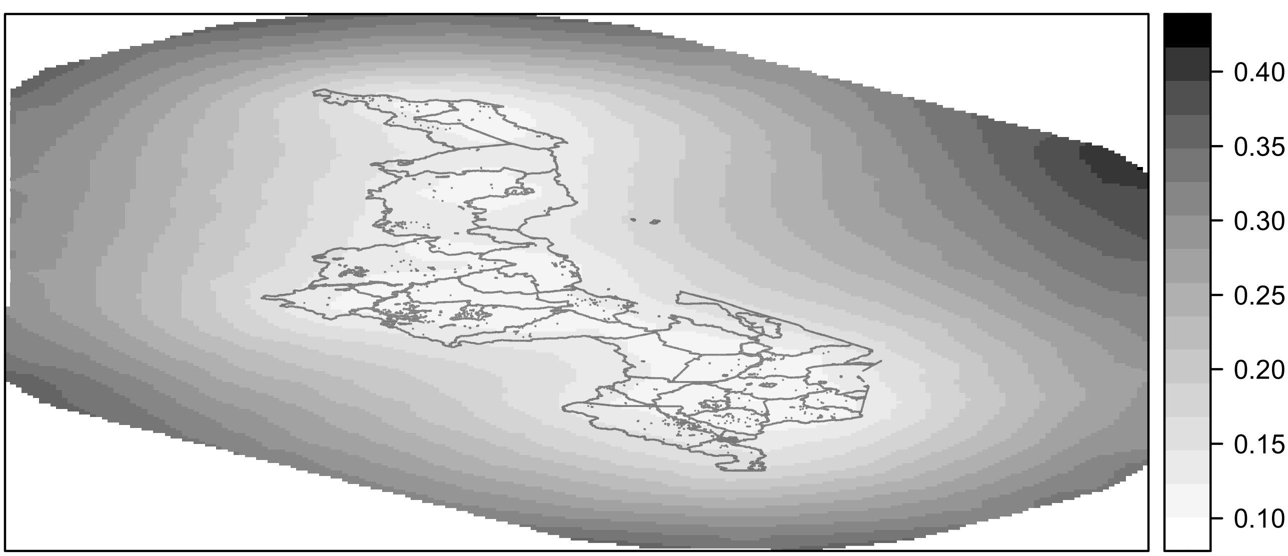


SD

buffer20

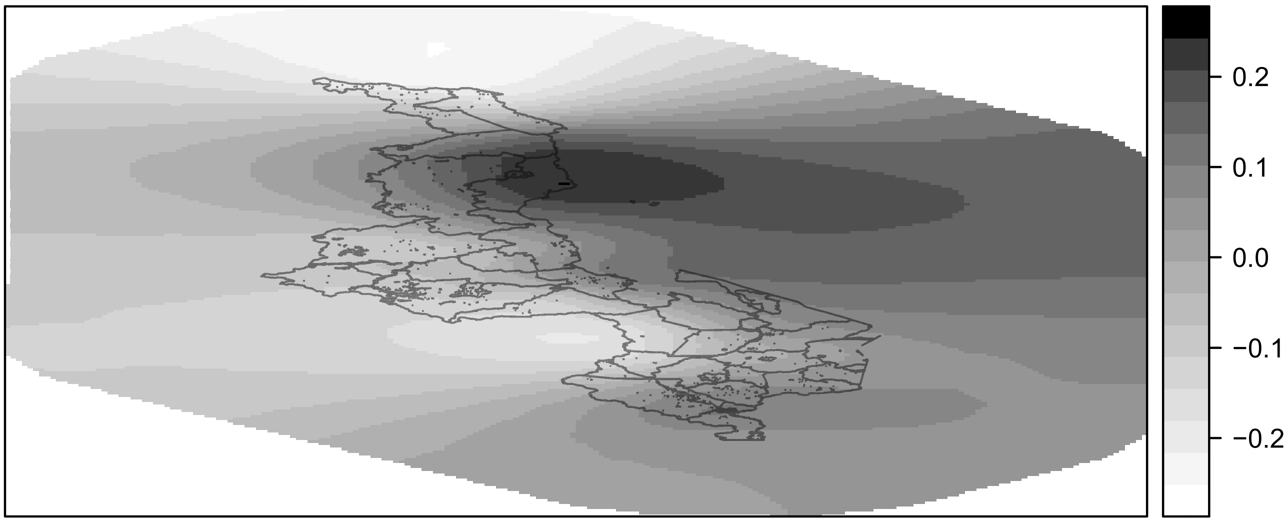


Mean

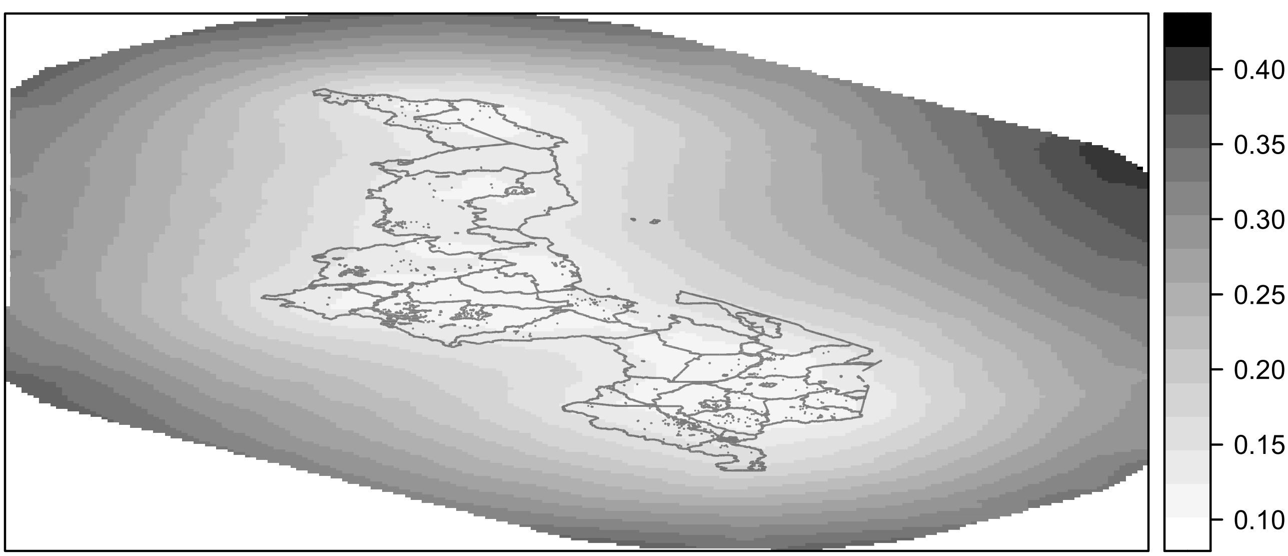


SD

buffer25

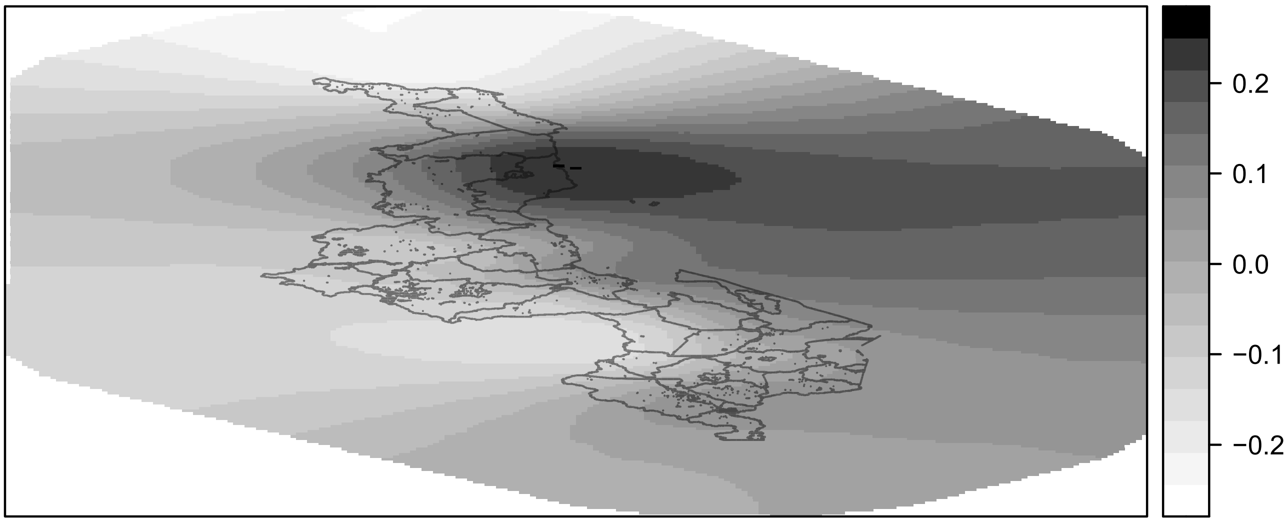


Mean

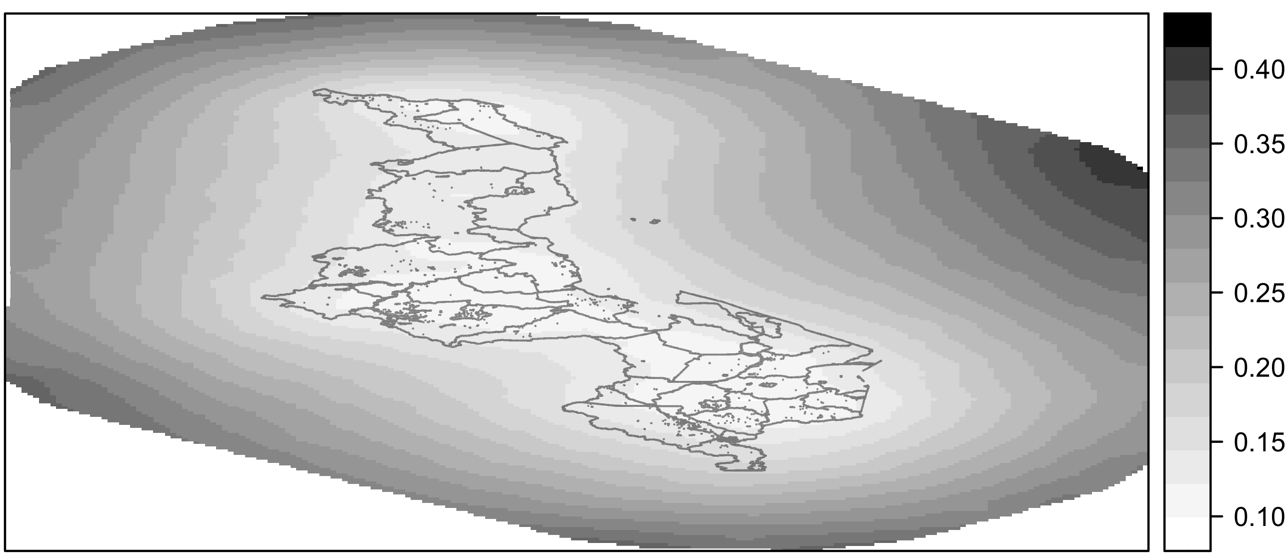


SD

buffer30

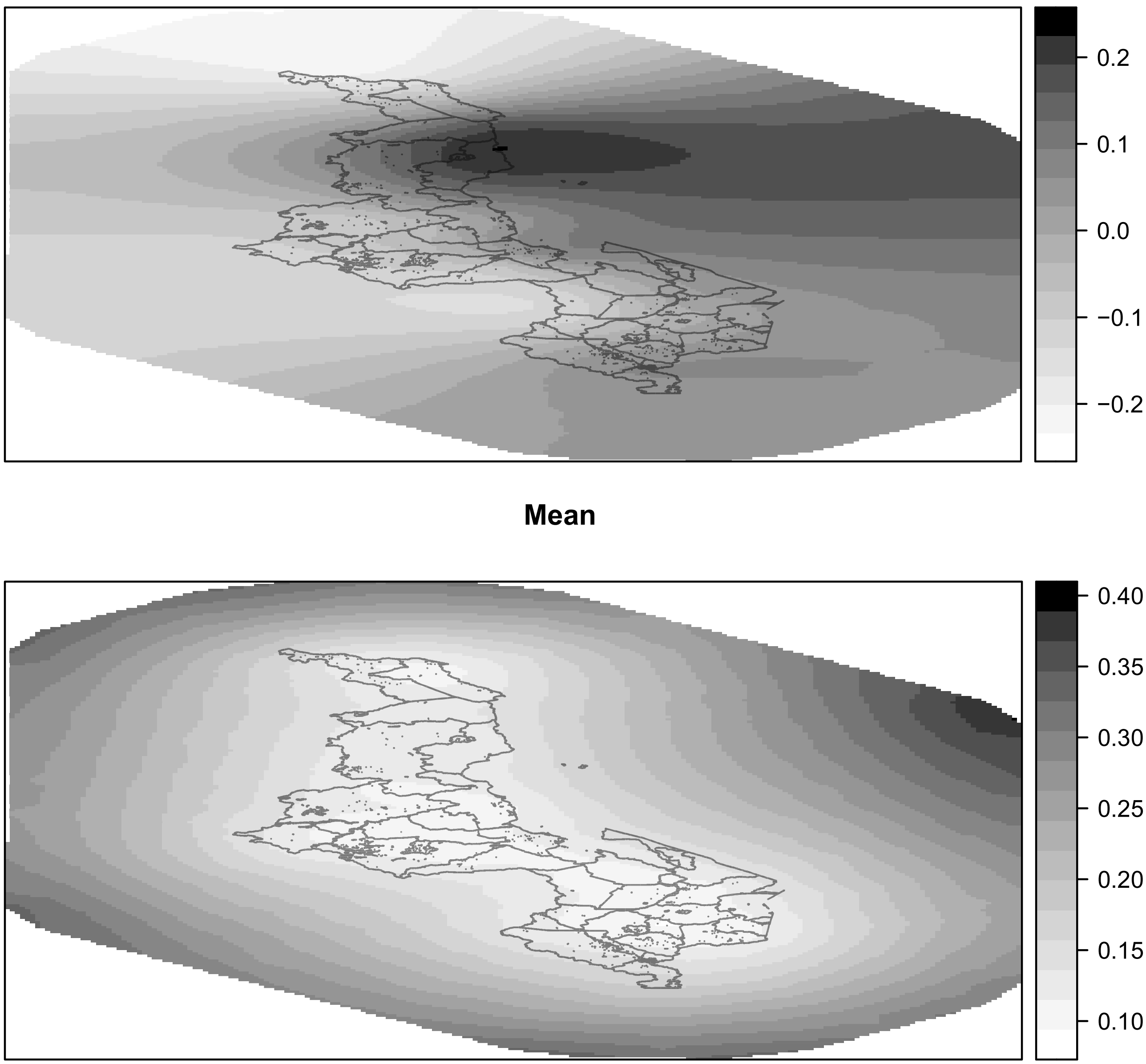


Mean

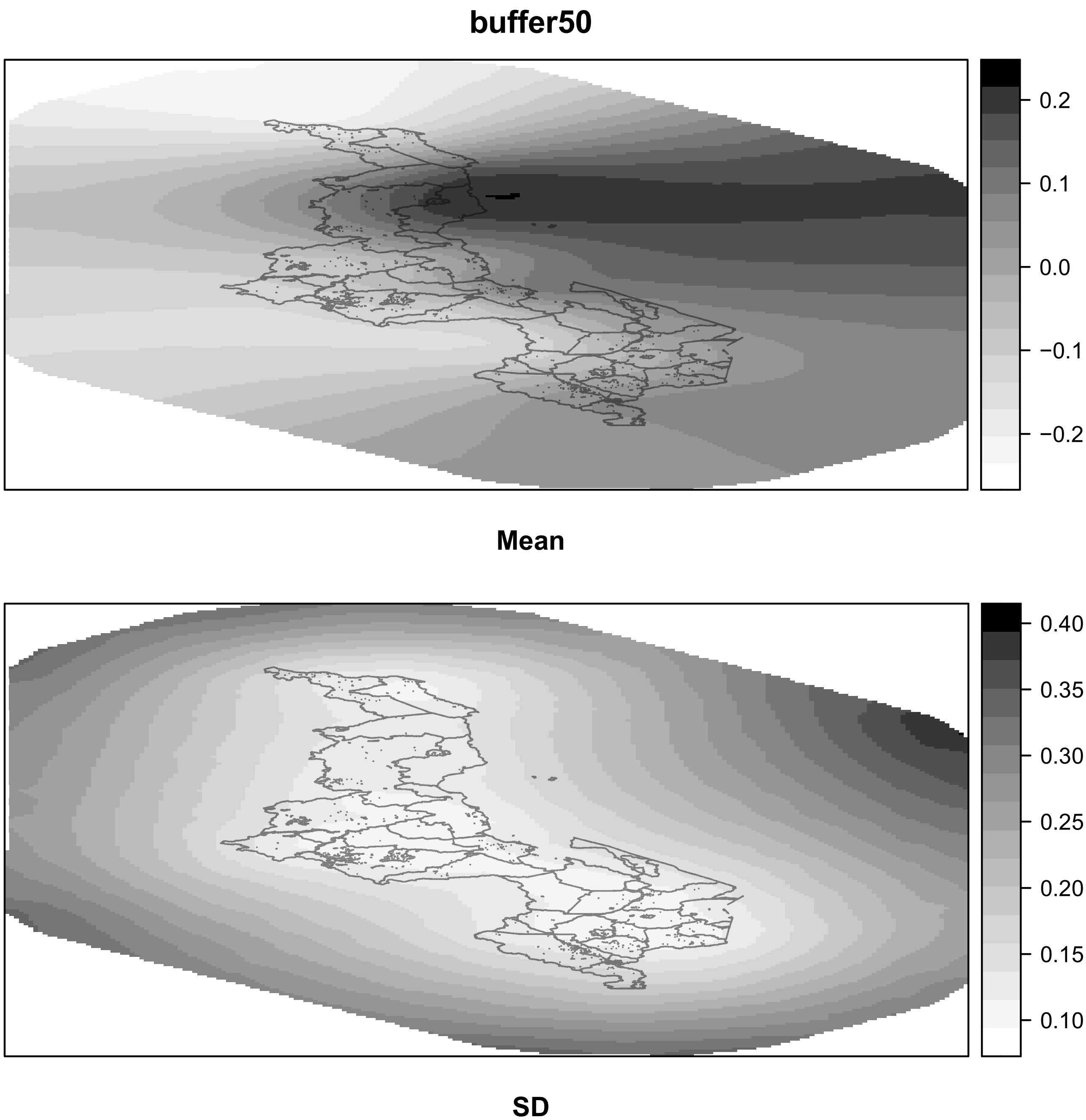


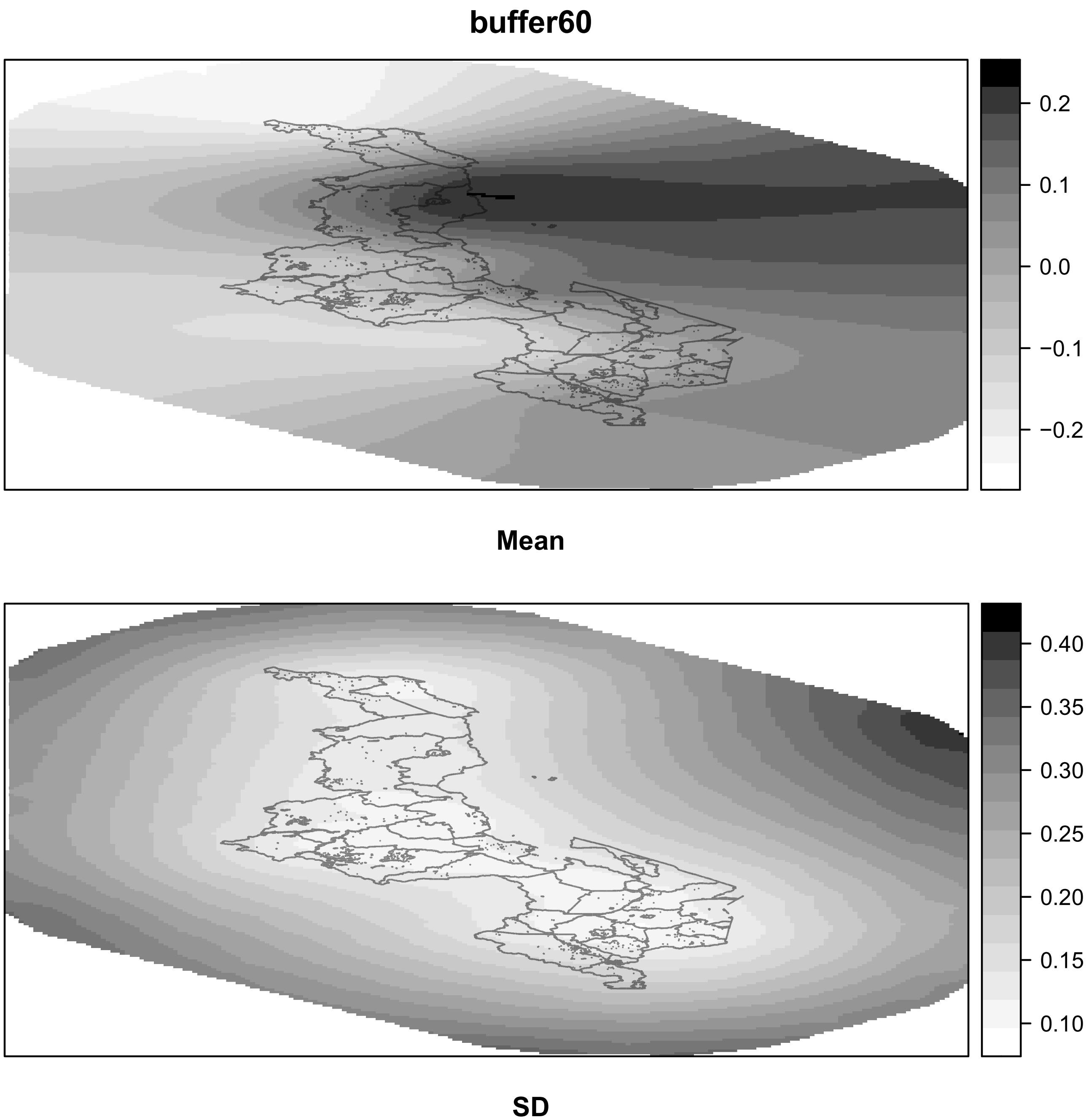
SD

buffer40

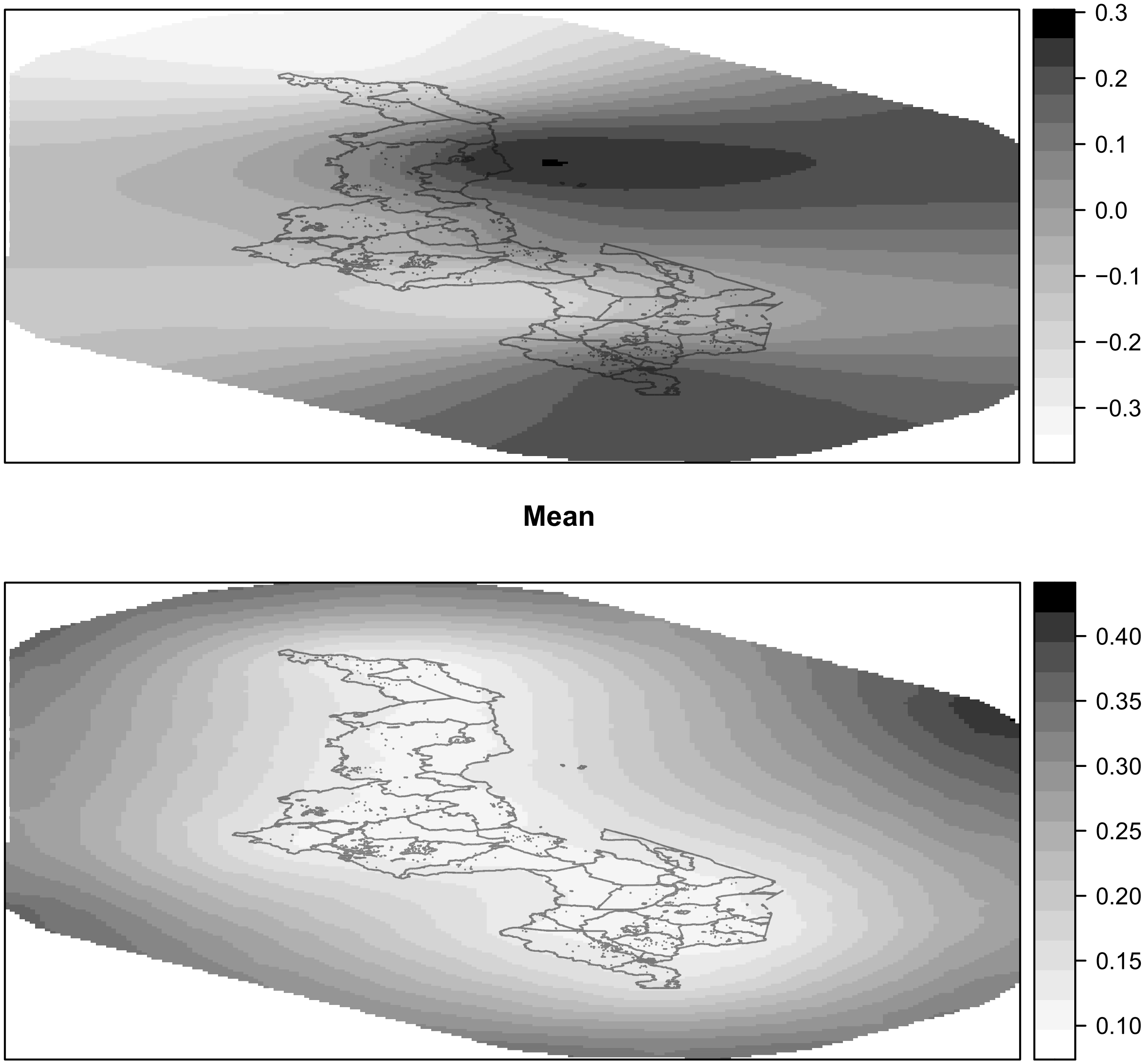


SD


buffer50
0.2
0.1
0.0
−0.1
−0.2
Mean
0.40
0.35
0.30
0.25
0.20
0.15
0.10
SD

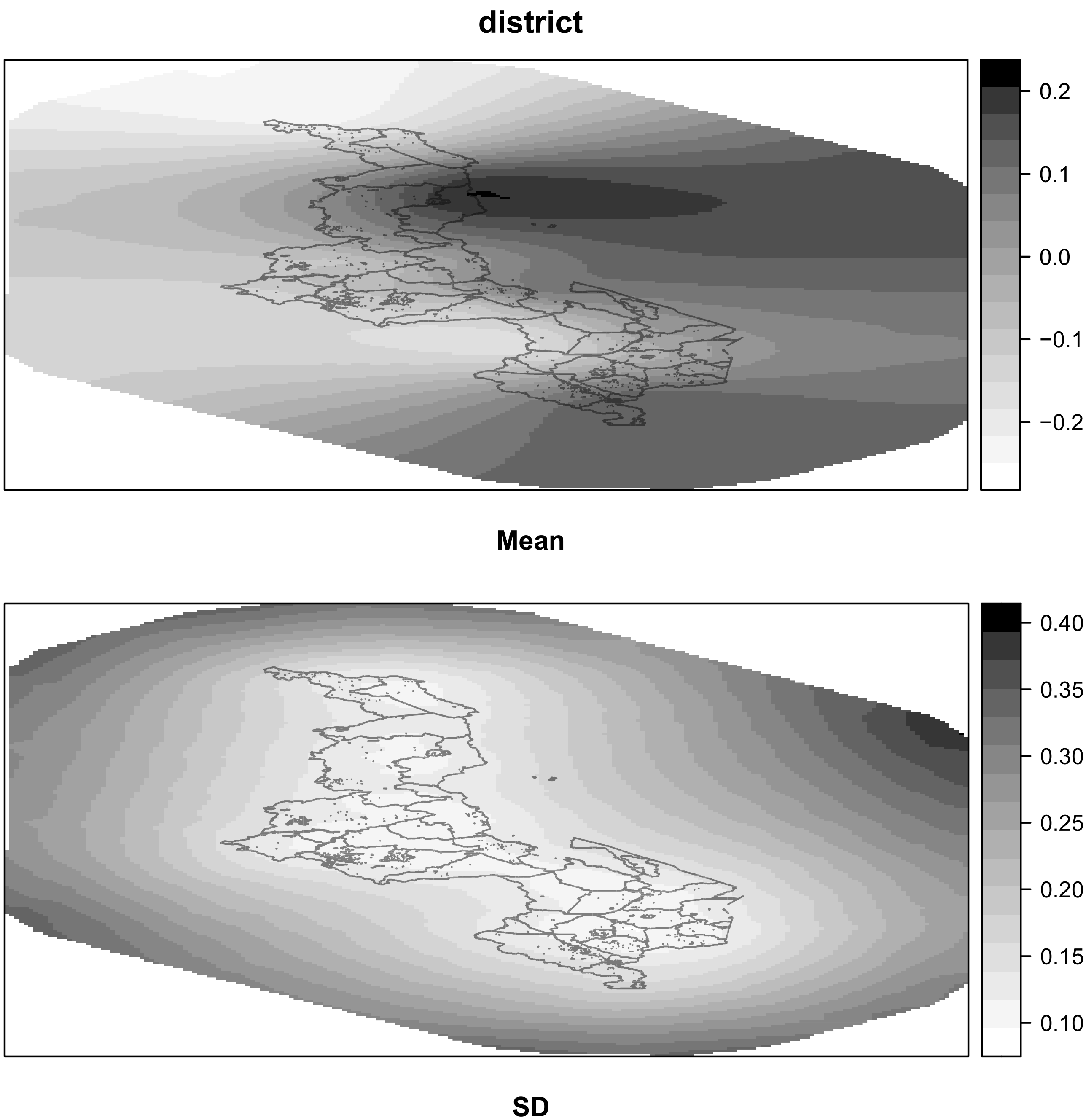

buffer60
0.2
0.1
0.0
−0.1
−0.2
Mean
0.40
0.35
0.30
0.25
0.20
0.15
0.10
SD

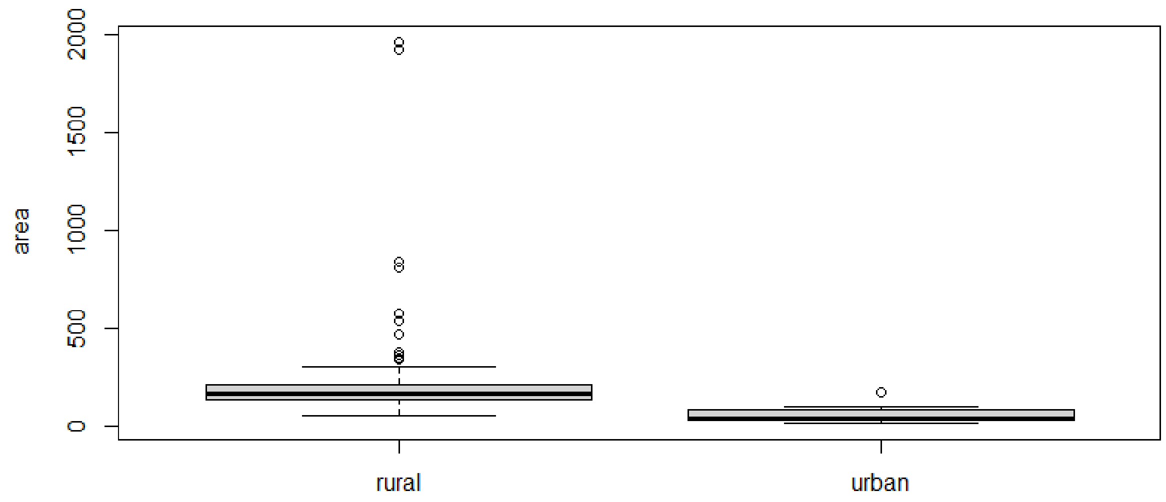
EA group



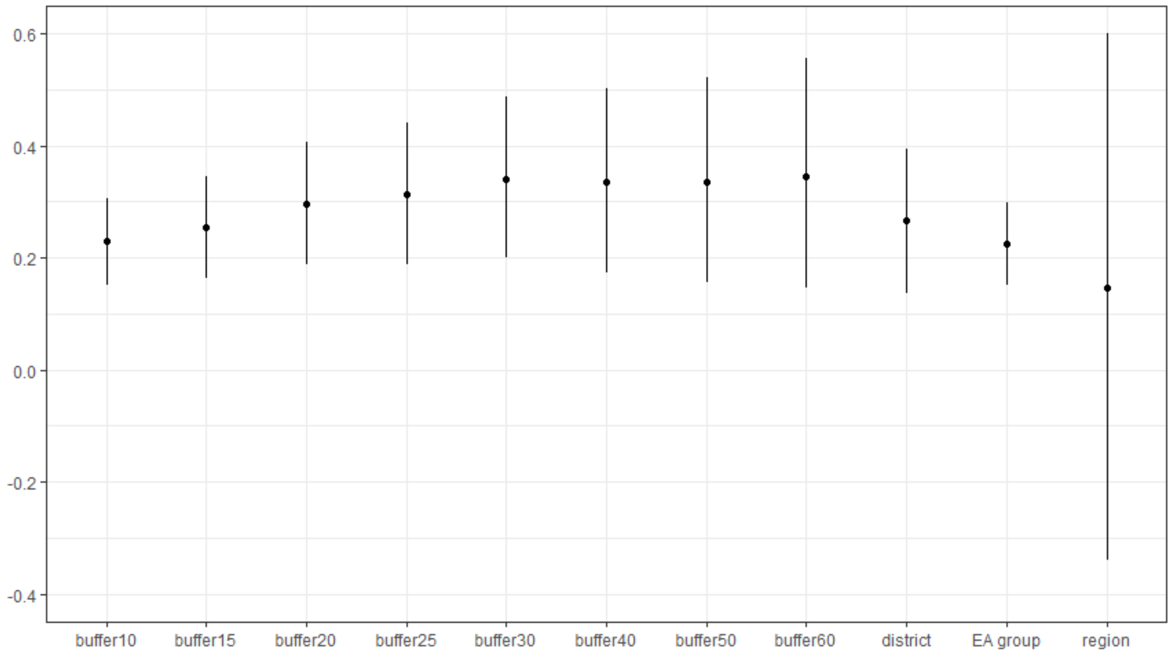
Mean

SD


district
0.2
0.1
0.0
−0.1
−0.2
Mean
0.40
0.35
0.30
0.25
0.20
0.15
0.10
SD



Supplementary Figure 7. Boxplots representing the area size (in km^2) of the EA group aggregation by rural and urban 2014-2015 MDHS cluster.



Supplementary Figure 8. Summary of the posterior distribution (the dot represents the mean and the bars 95% Credible Intervals) of the maize selenium aggregation (log-transformed) coefficient for all the ten models and the national-level model.

6 Chapter 6: Estimated risks of inadequate dietary selenium intakes in Malawi differ when using small area compared with national maize Se composition data

In this Chapter the implications of using small area maize Se composition for assessing dietary Se intake and adequacy risks were evaluated against current practices of using single average values, nationally and non-nationally produced.

The GitHub repository for replicating the analysis is here, and it builds on previous work done by the Micronutrient Action Policy Support projects. For instance, the cleaned version of the Malawi Fourth Integrated Household Survey, 2016-17 can be replicated using the *ihs4* repository and the Nutrient Conversion Table (NCT) used to assess dietary Se intakes (baseline NCT) can be reproduced using the *fct* repository.

Estimated risks of inadequate dietary selenium intakes in Malawi differ when using small area compared with national maize Se composition data

Abstract

The objective of this study was to assess the value of fine spatial resolution maize composition data for assessing population selenium intakes. To do so, dietary selenium apparent intakes and the prevalence of inadequacy across districts of Malawi were assessed under three scenarios: a) the baseline scenario, which represents 'typical' practice in nutrition assessment, used a single value of maize selenium composition from publicly available food composition tables, b) the national maize scenario, which used one single point national maize selenium composition data, and c) small area maize scenario used maize Se composition aggregated based on household Enumeration Areas. Both, the national and the small area maize scenarios, the maize selenium composition data was extracted from a high resolution (~250m^2) maize dataset. The results from the three scenarios were compared and interpreted in terms of potential relevance for public health nutrition policy.

Apparent dietary intakes of selenium were estimated using nationally representative household food consumption data from the 2016-17 Fourth Malawi Integrated Household Survey, which was matched to selenium composition data from public food composition tables and maize composition values depending on the scenario. Estimates of inadequate apparent intakes of selenium using the three maize composition scenarios were compared by district in Malawi.

The national prevalence of inadequate apparent intakes of selenium was substantially lower in the baseline scenario (40%, 95% CI 38-41) compared to the national (85%, 95% CI 83-85) or small area scenarios (79%, 95% CI 78-80). Furthermore, large differences in the prevalence of inadequate apparent intakes were observed in many districts for the single value scenarios and small area scenario. For example, in Nsanje district, the prevalence of inadequate dietary supplies was 74% (95% CI 66-82) when using a national level compared to 23% (95% CI, 12-35) when using small area maize data.

This study provides a clear illustration of the importance of using subnational food composition data for reliable population selenium nutrition surveillance. Our findings suggest that use of published FCT data would lead to substantial underestimation of selenium inadequacy risks in Malawi, while use of single national maize composition data would lead to large errors in estimates of deficiency risk at subnational scales. Thus, wider investment in subnational crop/food composition data – particularly for staple crops – is required to underpin population selenium nutrition surveillance and to ensure that policies are appropriately informed.

Introduction

Micronutrient deficiencies are widespread globally and underlie a large disease burden (Stevens et al., 2022). Estimates of dietary micronutrient intakes are important to help define the extent of micronutrient deficiencies and identify at-risk groups, and to guide potential interventions to alleviate deficiencies (Lividini & Masters, 2022).

Recently, the use of household surveys (e.g. Household Consumption and Expenditure Survey (HCES) data) has increasingly been used by the nutrition community as a source of dietary data to evaluate micronutrient apparent intake (Tang et al., 2022; World Health Organization & United Nations Children's Fund (UNICEF), 2022). To generate estimates of dietary micronutrient supplies, dietary data must be matched with food composition data, using a Nutrient Conversion Table (NCT). Information on the NCTs is usually sourced from Food Composition Tables and Databases (FCTs), however the limitations in current data availability and reporting practices (i.e. providing mean values of often unknown origin and lack of geographic information) precludes the possibility of location-informed matching of food composition data to consumption data, which may lead to inaccurate estimates of dietary micronutrients supplies (Segovia de la Revilla et al., 2023). This inaccuracy will be particularly important for mineral micronutrients with strong spatial variation in composition, for example selenium (Se), and for which lack of data in FCTs has been previously reported (Gashu et al., 2021; Segovia de la Revilla et al., 2023).

In Malawi, previous studies have used the Malawi Integrated Household Survey (IHS), a form of HCES, to evaluate the dietary micronutrient supplies of the Malawian population (Ecker and Qaim, 2011; Joy, Kumssa et al., 2015; Tang et al., 2021). Besides the food consumption module, the IHS also collects a wide range of other information, including household socioeconomic characteristics and geographic location of the household allowing for spatial analysis.

The main factors affecting dietary Se status are food intake and geographic origin of the food. Hence, in contexts with localised food systems and limited dietary diversity, information on household residency paired with small area Se food composition may perform better than national single values to identify populations potentially at risk of inadequate dietary Se intakes. Small area, which here refers to Enumeration Area level, Se composition data would be particularly important for the staple crop, maize, which is widely consumed in Malawi across all socioeconomic strata representing more than 60% of the overall energy supply. Maize is also reported to contribute approximately 20% of dietary Se intakes (Chilimba et al., 2011; Ecker and Qaim, 2011; Hurst et al., 2013; Joy, Kumssa et al., 2015). In addition to maize, fish is another important source of Se in Malawian diets, however, fish is unevenly consumed by the population in Malawi, with greater intakes in wealthier and urban households and households living close to Lake Malawi (Joy et al., 2015; Ligowe et al., 2020).

Several limitations on the use of FCTs for determining apparent Se intakes have been reported. For instance, in the Malawi FCT (2019) (van Graan et al., 2019), more than 70% of the Se concentration in the foods reported were missing. Hence, researchers and practitioners estimating Se supplies from diets in Malawi may turn to the Kenya FCT (2018) (FAO and Government of Kenya, 2018) or other national or regional FCTs in which maize Se composition may poorly reflect the reality in Malawi. Furthermore, the concentration of Se and other micronutrients in staple cereals are spatially variable (Gashu et al., 2021). Spatial variation in composition is highly relevant for population micronutrient nutrition, particularly in contexts with localised food systems. Thus, the use of subnational Se concentration in maize may produce more accurate estimates of dietary Se supplies, and we hypothesize that by combining geolocated household consumption data and small area Se concentration in maize grain, we can identify hotspots of populations at high risk of inadequate Se intakes in Malawi.

Unlike Malawi (and Ethiopia), where the GeoNutrition surveys were conducted, the majority of countries globally do not have large-scale surveys of crop/food Se composition, and/or these data are not available for nutrition researchers at subnational scales. This raises two important questions for nutrition surveillance: (1) In the absence of national sampled data, to what extent would borrowing food composition data from other countries lead to inaccurate assessment of dietary Se supplies and risks of inadequacy? And (2), how important is it to generate and use subnational Se composition data for key foods in localised food systems, rather than robust composition estimates at a national level?

Thus, the objectives of the current study were:

1. To estimate the dietary apparent Se intakes using composition data from published FCTs and to compare this to estimates using national and small area maize composition data.
2. To identify, compare and contrast estimates of Se inadequacy risk between districts in Malawi when using national average and small area maize composition data.

Methods

Food consumption data and household characteristics

Data on food consumption, socioeconomic characteristics, and geographic location of households were obtained from the Malawi Fourth Integrated Household Survey, 2016-17 (Malawi 2016-17 IHS4) which is run by the National Statistics Office (NSO) in Malawi with support from the World Bank Living Standards Measurement Study (LSMS). The sampling design, questionnaire, and data collection methods are described elsewhere (National Statistical Office, 2017). Briefly, the Malawi 2016-17 IHS4 was a nationally representative survey, in which participants were selected using a two-stage stratified (by urban/rural and districts) sampling design. In each stratum, enumeration areas (EAs) were selected using probability proportional to size based on the 2018 Malawi Population and Housing Census, and 16 households were randomly selected within each EA resulting in a sample size of 12,447 households, after accounting for 33 lost household interviews due to technical difficulties. Data were electronically collected by the enumerators from April 2016 to April 2017. Four modules from the household questionnaire were used in this study: Module A contained the household identification, survey information (e.g. sample weight) and household location (e.g. EA code). Module B contained the Household Roster with module D providing further demographic information on household members, including gender, age and whether women were pregnant. Module G reports food consumption over the past seven days at the household level using a standard list of food items. Additionally, the household identifiers and its displaced coordinates (i.e. modified latitude and longitude) were obtained from the Household Geovariates dataset which was prepared by the Malawi NSO (National Statistical Office, 2017).

The methods of processing household food consumption (7 days recall) data were reported previously (Mlambo et al., 2024; Tang et al., 2021). Following conversion of non-standard items to metric equivalents, food consumption at the household level was calculated and then allocated among household members per day proportional to their energy requirements using the Adult Female Equivalent approach (Weisell and Dop, 2012). The metric of apparent food consumption per AFE allows for comparison across households with varying demographic composition. Implausible consumption values were identified (e.g. maize flour > 2,000 g AFE^{-1} day^{-1} or > 5 log-SD), and replaced with median consumption of the entire sample. The dataset was cleaned using the scripts published in this repository.

Food composition data

The information on energy, water and Se content of all the foods reported in the Malawi 2016-17 IHS4 were extracted from the following publicly available FCTs: Malawi FCT (2019) (van Graan et al., 2019), Kenya FCT (2018) (FAO and Government of Kenya, 2018), Lesotho FCT (2006) (Lephole et al., 2006), UK FCT (Public Health England, 2021), USDA FCT (2019) (United States Department of Agriculture (USDA), 2019), in that order of preference. The FCTs were selected according to the relevance for the study and following the guidelines from the FAO/INFOODS (Charrondière et al., 2023). Food matches between the food list (n=129) reported in the Malawi 2016-17 IHS4 and the NCTs were performed following the FAO/INFOODS guidelines and the matches are provided in the Suppl. Table 1 and in the GitHub repository (FAO/INFOODS, 2012).

Georeferenced Se concentration in maize grain in Malawi

The small area and national (single-value) concentrations of Se in maize grains, for Malawi, were obtained from high spatial resolution (~250 m^2) estimates produced in the previous Chapter 5 and based on previous studies which measured Se content of whole grain maize samples collected from different locations across Malawi (Chilimba et al., 2011; Gashu et al., 2021; Kumssa et al., 2022). The majority (95%, 1,580/1,666) of the georeferenced maize Se sampled values were obtained from the GeoNutrition survey (Kumssa et al., 2022) and further details on the sampling plan, data collection and methods of Se analysis have been published elsewhere (Gashu et al., 2021; Kumssa et al., 2022). Briefly, paired soil-grain samples were collected from spatially balanced sampling locations across cropland areas of Malawi during the April-June 2018 harvest season. The remaining 5% (n=86/1,666) of data points were derived from Chilimba and colleagues (2011) who collected paired maize-soil samples at 73 sites (May 2009) covering 27 Extension Planning Areas (EPAs). The samples were collected at six Research Stations and 67 sites at farmers' fields representing most of the soil types under maize cultivation. In 2010, due to the high Se concentration found in the Shire Valley, a further 15 sites were sampled in that area. In both surveys, whole grain Se concentrations were measured by inductively coupled plasma-mass spectrometry (ICP–MS), and expressed as µg kg^{-1} dry weight.

The Nutrient Conversion Tables

Three NCTs were developed to estimate the apparent Se intakes in Malawi, which were the: baseline NCT, the national average maize NCT and the small area maize NCT (Table 1).

The baseline NCT extracted the Se content from relevant publicly available FCTs described above and was collated following the framework for reproducible, reusable, efficient and transparent NCTs

(Segovia de la Revilla et al., 2025). Further details regarding food matching have been published elsewhere (Tang et al., 2021) and can be found in the protocol and repository published elsewhere.

The national maize NCT used the same Se values as the baseline NCT for all foods, except maize and maize flours. The Se content of maize and maize flours were obtained from the high spatial resolution (~250 m^2) estimates produced in Chapter 5 and described in the section above ('Georeferenced Se concentration in maize grain in Malawi'). A single point of whole grain maize Se concentration was calculated as the national median Se value, and expressed as µg 100 g^{-1} of fresh weight edible portion (e.g. from 0.0280 mg kg^{-1} dry weight to 2.42 µg 100 g^{-1} fresh weight edible portion). National values (i.e. single points) of maize flour Se concentrations were calculated by taking the Se concentration of whole grain maize and adjusting to account for milling losses for 65% extraction, 95% extraction and bran maize flours, as done previously (Joy et al, 2015).

A small area maize NCT was constructed using a similar approach as the national maize NCT, however the concentration of maize and maize flours were estimated at cluster level (e.g. a group of EAs). The resulting NCT contained single Se concentration values to maize grain and each of the three maize flour fractions (after accounting for the milling losses) for each cluster location in the Malawi 2016-17 IHS4.

All the details and documentation to reproduce the three NCTs can be found here.

Calculating apparent Se consumption and inadequacy

The apparent intakes of Se and risks of inadequacy were assessed by combining information on apparent food consumption with the three NCTs. The apparent food consumption (in g per AFE per day) and the NCTs were combined to calculate apparent Se intakes. The prevalence of the population 'at-risk' of apparent inadequate Se intakes was calculated, using the Estimated Average Requirement cut-point approach (Carriquiry, 1999) and assuming an average Se requirement among adult women of 45 µg day^{-1} (Allen et al., 2020). The calculations were conducted for the three NCTs (the baseline NCT, the national maize NCT, and the small area maize NCT) and the results were compared.

Table 1. Summary of the Se food composition used in each of the three NCTs used to estimate dietary Se apparent intakes under the three scenarios.

NCT name	Spatial resolution	Summary of Se data sources	Maize items (food id) and Se values ref.	Notes
Baseline NCT	National	Kenya FCT (2018) (53%, n=68) Malawi FCT (2019) (33%, n =42) USDA FCT (2019) (9%, n=12) UK FCT (2021) (5%, n=6) Lesotho FCT (2016) (1%, n=1)	Maize ufa mgaiwa (normal flour) (101) – KE18(1022) Maize ufa refined (fine flour) (102) - KE18(1022) – 85% extraction Maize ufa madeya (bran flour) (103) KE18(1022) Maize grain (not as ufa) (104) – MW19(MW01_0037) Green maize (105) – KE18(1023) Maize – boiled or roasted (vendor) (820) – KE18(1051)	Information on which food entries were used specifically for maize and maize flours. The extraction rate was obtained for FAO FCT (FAO, 1953).
National average maize NCT	National	Primary data from Malawi (Chapter 5) (4%, n=6) Kenya FCT (2018) (48%, n=62) Malawi FCT (2019) (33%, n =42) USDA FCT (2019) (9%, n=12) UK FCT (2021) (5%, n=6) Lesotho FCT (2016) (1%, n=1))	Maize ufa mgaiwa (normal flour (101) - 100% extraction Maize ufa refined (fine flour) (102) – 65% extraction Maize ufa madeya (bran flour) (103) Maize grain (not as ufa) (104) Green maize (105) Maize – boiled or roasted (vendor) (820)	All Se values were the median Se concentration in the high-resolution maize dataset, adjusted for specific water content, and specific rate extraction for flours. The extraction rates were obtained from Joy et al., 2015.
Small area maize NCT	National except for maize and maize flours which were cluster level	Primary data from Malawi (Chapter 5) (4%, n=6) Kenya FCT (2018) (48%, n=64) Malawi FCT (2019) (33%, n =42) USDA FCT (2019) (9%, n=12) UK FCT (2021) (5%, n=6) Lesotho FCT (2016) (1%, n=1)	Maize ufa mgaiwa (normal flour (101) – 100% extraction Maize ufa refined (fine flour) (102) – 65% extraction Maize ufa madeya (bran flour) (103) Maize grain (not as ufa) (104) Green maize (105) Maize – boiled or roasted (vendor) (820)	All Se values were the median Se concentration in the high-resolution maize unique for each of the EA clusters in the Malawi 2016-17 IHS4. dataset, adjusted for specific water content, and specific rate extraction for flours. The extraction rates were obtained from Joy et al., 2015.

Results

The overall household characteristics of the Malawi 2016-17 IHS4 are presented in Table 2. In total 12,377 individual households were included in the analysis, after removing household without GPS location and/or EA identifier, and those with apparent energy intakes below 400 or above 9,000 kcal AFE^{-1} day^{-1} (Joy et al., 2015). In terms of apparent food consumption, most of the households reported consuming: 'vegetables, other and products' (98%), 'spices, other and products' (97%) and 'maize and products' (97%). The median apparent energy intake was 2,047 (IQR 1,484-2,836) kcal per AFE per day, of which 66% was supplied by 'maize and products', followed by 'sorghum and products' and 'cassava and products' (24% and 12%, respectively) (Suppl. Table 3). Notably, although 'sorghum and products' supplied high amounts of energy to households who reported its consumption, this was only 5% of the population.

Table 2. Summary statistics of the Malawi 2016-2017 IHS4. The table presents the percentage of the population living in rural and urban areas, and in the three regions of Malawi. In addition, it presents the percentage of households reporting consuming at least one of the six maize and maize flours reported in the Malawi 2016-2017 IHS4.

	Population % (N)		**Maize consumption** % (N)		**Maize consumption** Median (IQR) Kcal AFE^{-1} day^{-1}	
Urban	19	(2,244)	99	(2,202)	1,228	(883-1,708)
Rural	81	(10,104)	98	(9,831)	1,347	(962-1918)
Northern	9	(2,472)	93	(2,315)	1,272	(865-1,843)
Central	44	(4,191)	99	(4,165)	1,386	(1,016-1,932)
Southern	46	(5,685)	98	(5,553)	1,318	(911-1,907)

Most of the population (>93%) reported consuming at least one type of maize and maize flours across rural and urban residencies and all regions (Table 2). The median apparent consumption of maize and products was 364 g AFE^{-1} day^{-1} (IQR 257-518) with a slightly higher apparent consumption in rural than urban areas, i.e. 407 g AFE^{-1} day^{-1} (IQR 286-580) and 357 g AFE^{-1} day^{-1} (IQR 255-498), respectively. Similar patterns were found across all regions and districts in Malawi. The greatest median apparent maize and products consumption was observed in Central region (405 g AFE^{-1} day^{-1}, IQR 294-568), followed by Southern (391 g AFE^{-1} day^{-1}, IQR 267-573) and Northern regions (369 g AFE^{-1} day^{-1}, IQR 250-535). Among districts, the lowest apparent maize and products consumption was found in Nkhata bay (median 290 g AFE^{-1} day^{-1}, IQR 172-455) while the highest was in Mchinji (445 g AFE^{-1} day^{-1}, IQR 322-641).

In the baseline NCT (Suppl. Table 1), the majority of the Se values for matching to the Malawi 2016-17 IHS4 food list (n=129 food items) were from Kenya FCT (2018) (53%, n=68), followed by Malawi FCT

(33%, n=42), whereas the rest were derived from a combination of other FCTs (Table 1). Due to the scarcity of Se values in Malawi FCT (2018), all the Se values in maize and maize flours in the baseline NCT were extracted from the Kenya FCT (2018) except for 'maize grain (not as ufa)' which was obtained from the Malawi FCT (2019) (Table 1). The median and range of the georeferenced Se concentrations in maize used to calculate the Se composition of maize and maize flours in the national and small area NCTs are presented in Figure 1.

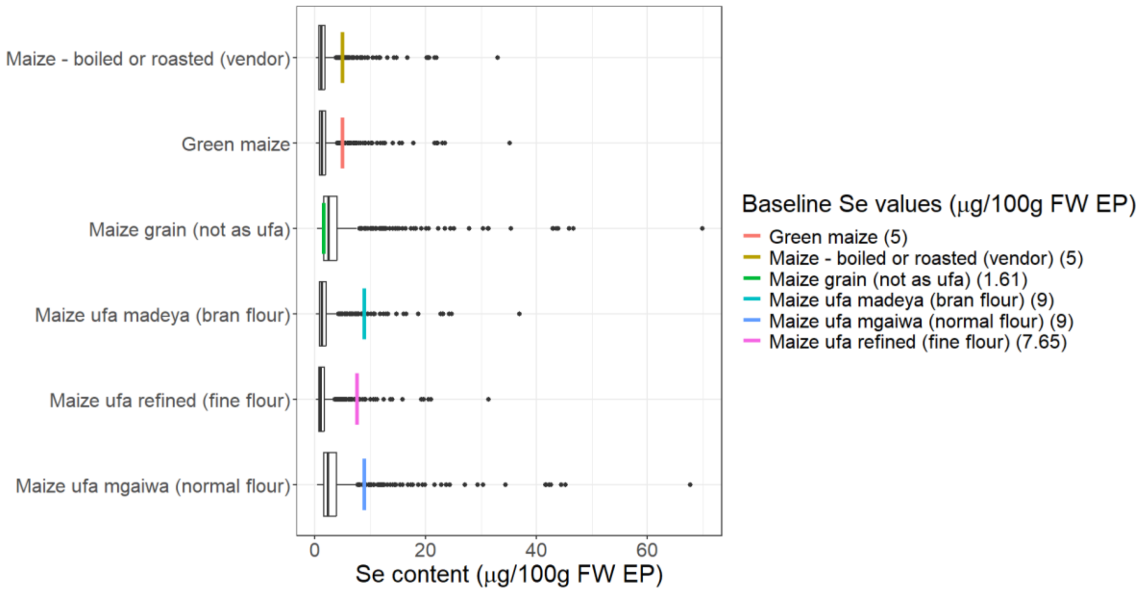


Figure 1. The selenium (Se) concentration in maize and maize flour used in the small area maize, national maize and baseline NCTs. The coloured-lines represent the Se content for those maize and flours in the baseline NCT, the boxplots represent the distribution of Se content in the small area maize NCT, and the black lines in the boxplots are the median Se content used in the national maize NCT. Fresh Weight Edible Product (FW EP).

Overall, the median Se concentration values for most maize items were higher in the baseline NCT than in the national and small area maize NCTs, based on the median and upper bound of the boxplots, as shown in Figure 1. Thus, the estimated median (IQR) apparent Se intake from the baseline NCT was the highest: 50.7 (36.7-70.3) µg AFE^{-1} day^{-1}, followed by the median (IQR) produced by small area maize Se values, 25.1 (15.7-40.3) µg AFE^{-1} day^{-1}. The lowest estimate of median (IQR) apparent Se intake was obtained when using the single national maize Se values, i.e. 23.1 (15.1-36.0) µg AFE^{-1} day^{-1}.

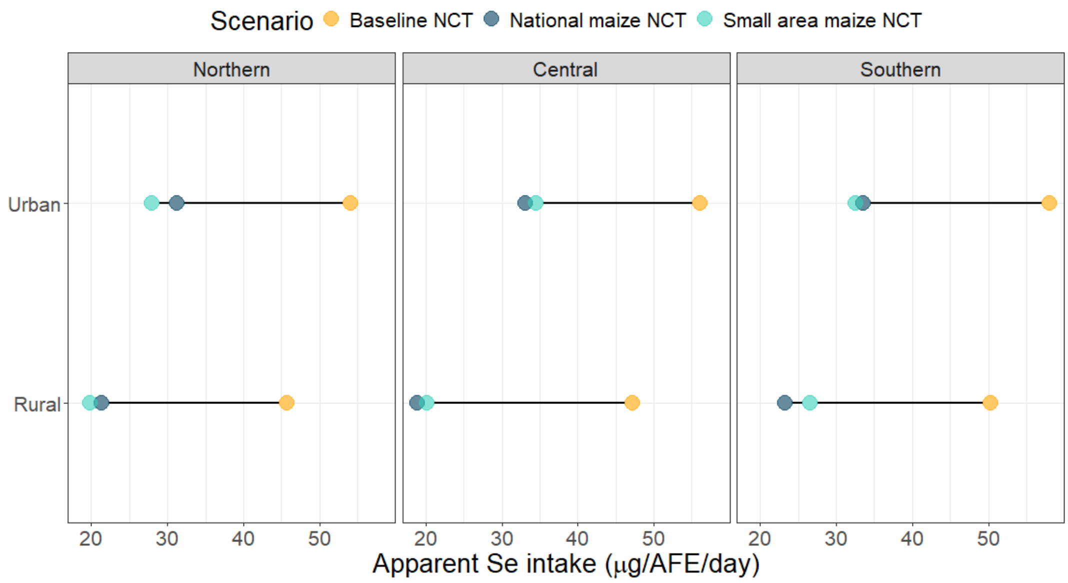


Figure 2: Comparison of the median apparent Se intakes (in µg AFE^{-1} day^{-1}) estimated using the baseline NCT (yellow), national maize NCT (dark blue) and small area maize NCT (light blue) by residency (urban/rural) and region in Malawi.

At the regional level, despite substantial differences in estimated Se intakes, the regional patterns observed when using baseline and small area maize NCTs were similar (Table 3), with lower median apparent intakes in the Northern and Central regions and higher median apparent intake in the Southern region. In contrast, when using the national maize NCT, median apparent Se intakes were equivalent in the Northern and Southern regions while the Central region had the lowest apparent Se intake (Table 3). Urban households had higher median apparent Se intakes than rural households in all three scenarios (Figure 2).

The majority of districts had higher median apparent Se intake when using the baseline NCT (Figure 3a), compared with other NCTs (Figure 3b and c), with the exception of two southern districts: Nsanje and Chikwawa. For these two districts the small area maize NCT generated the highest median (IQR) apparent Se intakes, with Nsanje showing the highest median (IQR) apparent intake of Se of all districts and NCTs, i.e. 78.9 (47.3-132.7) µg AFE^{-1} day^{-1}. In contrast, the use of the small area maize NCT generated the lowest median apparent Se intakes in Chitipa and Dedza compared with the other NCTs (Table 3 & Suppl. Fig. 1).

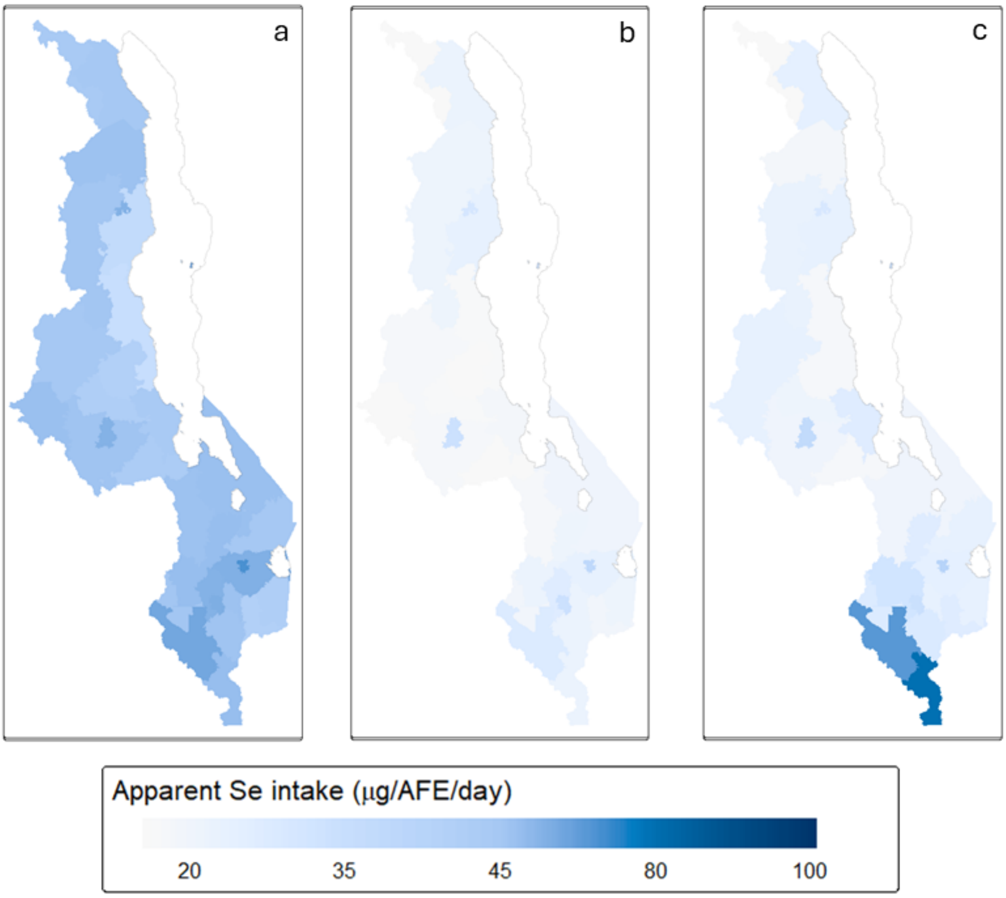


Figure 3. Comparison of the median apparent Se intakes (in µg/AFE/day) estimated using the baseline NCT (a), national maize NCT (b), and small area maize NCT (c) by district in Malawi. The areas with grey borders represent the three main lakes in Malawi: Lake Malawi, Lake Malombe and Lake Chilwa (from North to South).

The approach to defining maize Se composition resulted in large differences in the estimated prevalence of the Malawian population 'at-risk' of inadequate apparent Se intakes. The lowest percentage 'at-risk' was estimated using the baseline NCT, i.e. 40% (95% CI 38-41), followed by small area maize NCT (79%, 95% CI 78-80) and national maize NCT (85%, 95% CI 83-85). Furthermore, rural residents had higher estimated prevalence of apparent Se inadequacy independent of the NCT approach, i.e. 42% (95% CI 41-44), 82% (95% CI 81-84) and 87% (95% CI 86-88) among rural residents and 28% (95% CI 25-31), 67% (95% CI 64-71) and 68% (95% CI 65-71) among urban residents, for the baseline, small area maize and national maize NCTs, respectively. After disaggregating by region, comparable trends were seen (i.e. baseline estimates < small area estimates < national estimates), except that the estimated percentage of people 'at-risk' of apparent Se inadequacy in the Northern region of Malawi was highest when using the small area maize NCT instead of the national NCT (Suppl. Table 2).

Table 3. The apparent energy intakes and the three apparent intake of Se scenarios based on each NCT: Baseline NCT, National (maize Se) NCT, and small area (maize Se) NCT.

	Apparent Energy intake (Kcal/AFE/day)		*Baseline apparent Se intake*		*National apparent Se intake*		*Small area apparent Se intake*	
	Median (IQR)		Median (IQR) (µg AFE^{-1} day^{-1})					
Urban	2431	(1807-3261)	57.91	(42.78-80.01)	34.09	(22.81-50.84)	34.39	(22.4-51.85)
Rural	1956	(1421-2726)	48.95	(35.43-68.11)	21.36	(13.89-32.48)	23.19	(14.57-37.39)
Northern	2109	(1605-2889)	48.6	(34.18-69.72)	24.17	(15.54-38.1)	22.51	(13.89-36.03)
Central	2043	(1510-2789)	49.28	(36.64-66.22)	20.96	(13.93-32.88)	22.6	(14.36-35.83)
Southern	2034	(1436-2864)	52.56	(37.21-74.15)	25.13	(16.25-39.33)	28.38	(17.9-46.34)
Balaka	2060	(1479-2797)	52.41	(37.6-71.76)	23.43	(15.73-34.25)	27.1	(18.03-40.06)
Blantyre	2163	(1660-2870)	57.5	(42.41-73.27)	28.96	(19.94-41.94)	32.23	(20.98-46.31)
Blantyre City	2472	(1876-3204)	58.95	(43.72-78.8)	35.84	(23.65-52.39)	34.39	(22.1-51.36)
Chikwawa	2039	(1363-2982)	60.93	(37.93-88.15)	30.09	(15.79-58.89)	64.96	(40.32-105.24)
Chiradzulu	1859	(1353-2757)	50.75	(36.83-71.09)	24.27	(16.68-37.63)	23.01	(15.83-36.21)
Chitipa	2091	(1584-2813)	47.88	(35.57-65.25)	18.26	(12.68-27.5)	15.36	(9.08-24.87)
Dedza	1943	(1506-2605)	46.85	(35.83-60.68)	17.8	(12.14-25.27)	17.06	(10.52-24.66)
Dowa	1965	(1468-2603)	46.25	(36.18-64.12)	19.2	(12.61-30.64)	21.18	(14.35-34.13)
Karonga	2053	(1555-2703)	48.59	(31.43-69.58)	23.6	(15.73-41.26)	24.42	(15.6-43.96)
Kasungu	2020	(1469-2575)	48.38	(34.82-63.1)	20.07	(13.46-30.56)	23.34	(15.28-38.53)
Likoma	2204	(1635-2922)	61.39	(43.47-85.59)	35.02	(24.77-55.76)	35.02	(24.77-55.76)
Lilongwe	2047	(1544-2728)	50.25	(38.48-64.52)	19.99	(14.23-29.25)	20.31	(13.54-30.01)
Lilongwe City	2448	(1777-3381)	57.53	(44.56-79.75)	35.23	(23.92-51.42)	36.64	(26.43-53.53)
Machinga	1859	(1319-2612)	48.1	(33.95-64.85)	22.01	(14.55-31.46)	23.02	(14.13-34.08)
Mangochi	2068	(1482-2999)	51.88	(36.7-76.25)	22.34	(14.49-32.06)	20.37	(13.07-31.7)
Mchinji	2133	(1573-2958)	51.26	(36.64-71.66)	18.37	(11.77-32.34)	23.65	(14.87-38.24)
Mulanje	1672	(1164-2277)	45.12	(31.66-59.21)	21.24	(13.54-33.04)	26.93	(16.94-41.5)
Mwanza	1984	(1470-2701)	49.13	(34.28-64.82)	21.08	(14.09-33.44)	27.64	(16.77-42.57)
Mzimba	2127	(1652-2924)	49.35	(36.46-69.38)	22.74	(14.18-34.99)	22.71	(14.51-34.47)
Mzuzu City	2373	(1737-3207)	57.08	(39.68-79.38)	32.44	(23.54-52.92)	29.33	(19.73-48.78)
Neno	2181	(1540-2870)	52.15	(36.06-70.04)	23.89	(15.38-34.96)	31.66	(20.07-46.9)
Nkhatabay	1935	(1485-2660)	40.24	(24.36-64.41)	25.33	(15.9-36.79)	23.39	(14.9-34.85)
Nkhotakota	1777	(1300-2577)	38.46	(28.21-55.15)	18.87	(13.91-28.32)	17.65	(12.44-27.87)
Nsanje	1929	(1283-2744)	52.62	(37.34-77.31)	24.05	(14.95-46.13)	78.93	(47.26-132.67)
Ntcheu	2089	(1537-2922)	50.67	(36.69-69.37)	19.9	(13.61-31.32)	19.97	(13.07-30.89)
Ntchisi	1882	(1355-2515)	44.82	(32.56-60.03)	17.76	(11.08-27.28)	17.79	(10.83-28.37)
Phalombe	1583	(1096-2191)	46.15	(32.63-62.87)	22.7	(13.41-33.41)	23.28	(13.32-38.29)
Rumphi	2172	(1601-2967)	51.18	(37.62-69.18)	23	(14.85-34.54)	18.71	(11.37-30.19)
Salima	1799	(1316-2442)	47.73	(35.02-61.24)	20.76	(14.19-28.87)	26.8	(17.14-50.11)
Thyolo	1820	(1357-2802)	49.86	(34.19-74.62)	24.23	(15.4-37.07)	28.5	(18.35-47.2)
Zomba	2220	(1605-3047)	58.15	(42.58-74.46)	25.57	(18.35-38.57)	24.16	(16.92-37.47)
Zomba City	2833	(2051-3741)	69.06	(50.36-90.74)	39.64	(27.57-57.51)	40.65	(28.06-57.71)

Finally, when evaluating the apparent Se inadequacy at district level, as illustrated in the Figure 4, the lowest percentages 'at-risk' of inadequate intakes were found when using the baseline NCT for all districts with the exception of two districts in the South – Nsanje and Chikwawa – where the lowest percentages of Se inadequacy risks were estimated using the small area maize NCT (Figure 4). When comparing the small area and national maize NCTs, 59% (19/32) of the districts showed apparent Se inadequacies within 3%-point difference. For the other districts, except for Mzuzu city and Rumphi, the percentage of population at risk of inadequate Se intakes was lower using the small area NCT compared to the national maize NCT (Figure 4 and Suppl. Table 4).

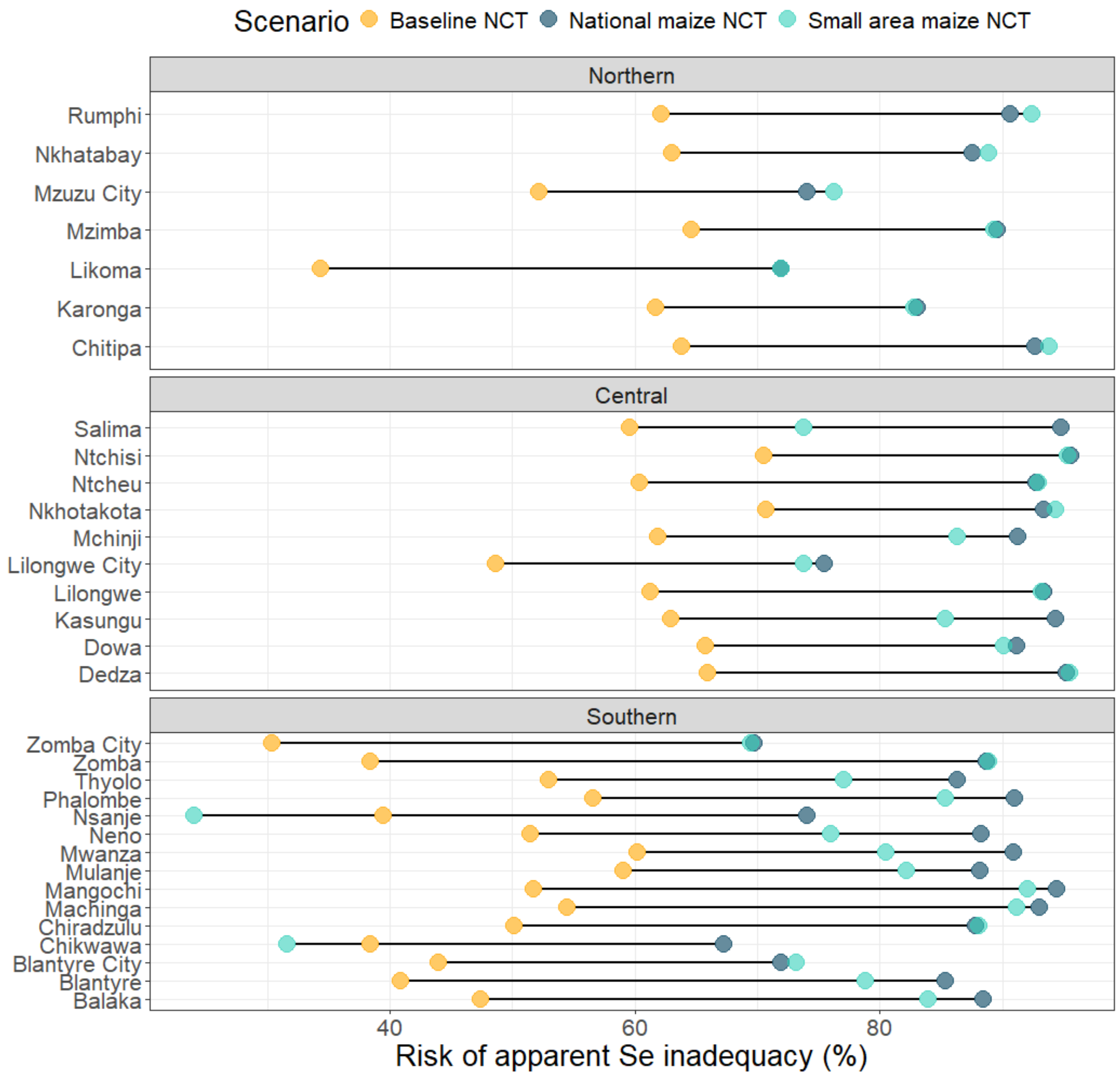


Figure 4. Percentage of population 'at-risk' of apparent Se inadequate intake using the three NCTs: baseline NCT (yellow), national maize NCT (dark blue) and small area maize NCT (light blue) by district and region in Malawi.

Discussion

Primary findings

This study provided a method to operationalise the spatial integration of mineral micronutrient composition data and dietary data to support accurate estimates of apparent dietary Se intakes and the prevalence of inadequate apparent intakes in Malawi. The study demonstrates the importance of spatial variation in maize Se concentration for driving variation is dietary Se intakes and inadequacy risks, and the need for high spatial resolution Se data for accurate Se nutrition surveillance. In comparison to the small area data, the baseline NCT, which used published data on maize composition mostly from non-Malawian samples, led to a large under-estimation of the prevalence of Se inadequacy at national and subnational levels, with the exception of two southern districts in Malawi: Nsanje and Chikwawa. The national maize NCT, which used composition data from Malawi samples but applied these as an average across districts, generated a similar estimate of the prevalence of inadequate Se intakes at national level, but failed to capture important subnational variation, for example leading to an over-estimation of the prevalence of Se inadequacy in Nsanje, Chikwawa, Salima and Neno districts.

To our knowledge, this is the first time that small area Se concentration of maize for the whole country is used (i.e. one Se composition value per cluster) and paired with nationally representative apparent food consumption data based on their geographic location (e.g. cluster GPS location). This method provided cluster specific maize values for each household in the Malawi 2016-2017 IHS4 to capture spatial variations in maize Se content and therefore apparent Se intakes. The results were evaluated against current practices in nutritional assessment, e.g. the use of single point FCTs values to compile a survey-specific NCT, and the use of single-national maize values instead of small area values. Overall, the use of the small area maize NCT provided different estimates of both apparent Se intakes and the percentage of the population at risk of inadequacy. The differences were particularly relevant when disaggregating the data at subnational scales, because it allowed the identification of districts at high and low risk of inadequacy, where the observations would be masked by the use of single point NCTs (i.e. baseline and national maize NCT). Furthermore, these variations in estimates would not reflect substantial subnational differences in apparent maize and maize flour consumption, as patterns across rural/urban, regions and districts were generally similar. Even when comparing the districts with the highest and lowest apparent maize and maize flour consumption, such as Mzuzu City and Mchinji, the differences were relatively small.

The baseline NCT provided the highest apparent intake estimates of Se, and in turn, the lowest prevalence of apparent Se inadequacy, because the Se content of all the maize and maize flours (excl. 'maize grain (not as ufa)') were collated from the Kenya FCT (2018). The Se concentration of maize in Kenya is higher than in Malawi, according to Ngigi et al. (2020), due to inter-country differences in soil types and other environmental characteristics. Moreover, the actual location of the Se values published in Kenya FCT (2018) is unknown (Segovia de la Revilla et al., 2023). These issues highlight concerns previously raised regarding the appropriateness of using values from Kenya or other neighbouring countries, when assessing Se intakes particularly for highly consumed foods (Combs, 2015). As shown in Figure 4, the overall prevalence of inadequate dietary Se status in Malawi would appear low when using the baseline NCT, which may lead to inappropriate nutrition-related policy decisions.

In contrast, the national maize NCT, which used one single value from the high-resolution maize Se estimates, provided the lowest apparent Se intakes and the highest apparent Se inadequacy. This result reflects the predominance of low plant-available Se in soils in Malawi and, thus, the small concentrations in maize (and other crops), which influenced the median values used, for maize and maize flours, in the national NCT for Malawi (Ligowe et al., 2020). High Se areas were sparse and only found in certain part of the country, such as the Shire Valley, where pockets of calcareous soils were contributing to increased uptake of Se into plants grown in those areas. Hence, the use of single-national maize Se values instead of small area data, particularly in these areas, would mask higher apparent intakes as illustrated by the results for Nsanje and Chikwawa.

Finally, the use of high-resolution estimates of crop composition data paired with small area estimation methods, could enable further disaggregation of the estimates of dietary intakes and inadequacies which in turn, would support the identification of vulnerable populations. For instance, in areas of very low maize Se concentration where rural households from the least wealthy quintile may be more at risk than their wealthier and urban counterparts. In contrast, resources allocated to alleviate Se inadequacies in areas with high Se in maize misidentified by national-level single values (e.g. Nsanje) may be relocated to those areas or populations that may be more in need.

Comparison with previous literature

Our findings are consistent with previous studies that revealed the importance of small area mineral composition data to estimate dietary intakes of minerals at sub-national scales (e.g. districts or counties) in Tanzania, Kenya and Malawi (Watts et al., 2015, 2019), and studies that considered the dietary contribution of particular crops (e.g. cereals) in Ethiopia and Malawi (Abdu et al., 2022; Broadley et al., 2012; Chilimba et al., 2011). Previously, Joy, Kumssa et al. (2015) estimated dietary mineral intakes and inadequacy risks in Malawi through integration of household survey data and georeferenced mineral sampled data aggregated by broadly grouped soil type. Despite the limited composition data generated through convenience sampling rather than a designed survey such as the GeoNutrition survey, it showed the importance of spatial variation and the small area composition data, not only for estimating apparent Se intakes, but also for Ca and Zn.

Hurst et al. (2013) reported 10-fold difference in maize Se concentration between northern (Zombwe Extension Planning Area (EPA), Mzuzu district) and southern (Mikalango EPA, Chikwawa district) villages in Malawi. This, in turn, led to higher dietary Se and plasma Se concentration in women (18-50 years old) living in the southern district of Chikwawa than in the northern district of Mzuzu. Despite differences in the absolute values reported in Hurst et al (2013) and in our study, the small area scenario was the only scenario able to capture differences in the dietary Se intakes between the two districts (Chikwawa and Mzuzu, Table 3).

Comparison with plasma Se in Malawi

In Malawi, 65% of women (15-49 years old) were at risk of biochemical Se deficiency (Phiri et al., 2019). These estimates align with the estimated prevalence of inadequate Se intakes, in our study, when using either the small area maize NCT and less so when the values of the national maize NCT or compiled NCT were used (prevalence: 79% (95% CI 78-80) vs 85% (95% CI 83-85) vs 40% (95% CI 38-41), respectively). Furthermore, when comparing the risk of biochemical Se deficiency for rural (66%, 95% CI 56-76) and urban women (15-49 years old) (55%, 95% CI 42-69), in Phiri et al. (2019), similar patterns were found in our study, particularly when the prevalence of population at risk of apparent inadequate Se intakes were estimated using the small area maize NCT. Interestingly, when comparing the risk of deficiency at regional level, although the differences between the three regions were captured by the use of small area maize NCT, it did not fully capture the magnitude of the reduced risk of deficiency seen in the Southern region of Malawi (57%, 95% CI 43-71).

The estimated prevalence of population at risk of inadequate apparent Se intakes, when using the small area and national maize NCTs, were, however, consistently higher than those found in the national micronutrient survey. There are several reasons that could account for this discrepancy. Firstly, the dietary data used did not measure individual-level food consumption as it was collected at household level, and secondly, these datasets are not considered as accurate as other methods for measuring food consumption (FAO & The World Bank, 2018; Gibson & Ferguson, 2008). Hence, our risk estimates of Se inadequacy may be affected by inaccuracies due to the data collection methods, including unequal intra-household food distribution and by underreporting of certain foods (Tang et al., 2021; Wheeler, 1991). Additionally, the survey used a standardised list of food items which could lead to the misreporting of foods consumed outside of the food list. An example could be the consumption of *gramil flour*, which has been reported as one of the preferred or key maize flour to make *nsima*, the main staple food in Malawi (Chiutsi-Phiri et al., 2021; Mlotha et al., 2016). This type of maize flour, which is only partially refined, was not specified in the food list in the Malawi 2016-17 IHS4, and its consumption may have been recorded within the item 'maize flour refined' or not recorded at all. Both scenarios would have contributed to the underestimation of apparent Se intakes, and subsequent overestimation of the apparent Se inadequate intakes. The values used to account for Se losses during milling might have over-estimated the actual losses due to differences in the maize degree of processing, crop location and variety (Ekpa et al., 2019; Gwirtz & Garcia-Casal, 2014). For instance, bran flour would have higher Se content than whole grain although we used whole grain Se values. On the other hand, differences in the data collection of the two surveys may have contributed to the discrepancies. For instance, the 2015-16 MDHS-MNS was collected during the 'lean season' which generally spans from October to February while the Malawi 2016-2017 IHS4 was collected throughout the year, from April 2016 to April 2017. In Malawi, there is a seasonal pattern in the types of maize flour consumed. The high Se content maize flour (i.e. 95% extraction flour) is consumed more frequently in the lean season than in the food plenty season (Ecker & Qaim, 2011), and this seasonal effect might contribute to the discrepancies. Finally, another reason, particularly in the Southern region, could be the contribution of other cereals which have higher Se concentration, and for which the lack of the small area Se data may have underestimated the true Se content. This is because the Southern region in Malawi is known for its calcareous soils and for their contribution to higher Se uptakes by plants where higher Se concentration have been often reported in maize as well as in other crops (Gashu et al., 2021; Joy et al., 2015).

Strengths and Limitations

The main strength of this study is the use of high-resolution maize Se estimates that allowed to generate small area location-specific maize Se concentration for use in the prevalence of apparent Se inadequacy for Malawian population. Additionally, this study used open science frameworks and scripted approaches which allow for further comparability, transparency and review of the methods. The repository containing all the step needed to produce these estimates could be adapted and re-used to produce similar results for other mineral micronutrients. For instance, a similar modelling analysis have assessed the impact of using district level maize Zn composition for estimating apparent Zn intakes in Malawi (Botoman et al, forthcoming) while another study could be developed for apparent Ca intakes which have previously shown important spatial variation (Gashu et al., 2021; Joy et al., 2015). Furthermore, this work could be expanded to other geographic areas, for instance, following the study from Abdu and colleagues (2022) in Ethiopia where they evaluated the spatial variation of district level mineral concentration of staple grains, these scripts could be adapted to evaluate their potential impact on apparent micronutrient intakes in Ethiopia.

There are some limitations to this study, for example, the georeferenced sampled data used was only available for maize, although the spatial variation in Se concentration in other crops would likely influence further the apparent Se intakes. Other limitations are related to the household survey which has been previously described (FAO & The World Bank, 2018; Moltedo et al., 2022; Tang et al., 2021). Firstly, the location of the households are displaced which may have affected the accuracy of the small area maize Se concentration. However, to minimise the displacement bias the maize Se concentration were aggregated in areas sizes that would cover all the displacement of the coordinates (More information can be found in Chapter 5 and in the GitHub repository). Furthermore, there are a number of issues regarding the consumption information collected in household surveys, including recall periods, acquisition (i.e. food reported as purchased not as consumed), food reported at household level and food consumed away from home among other. For instance, by reducing the upper energy constrains from 9,000 kcal $AFE^{-1}day^{-1}$ to 5,000 kcal $AFE^{-1}day^{-1}$, a proportional increase of ~1% in the estimates of the apparent inadequacy found in the three scenarios (baseline NCT: 41% (95% CI 40-42) vs 86% (95% CI 85-86) vs 81% (95% CI 80-82), respectively. Despite these caveats, which have been previously reported, the bias introduced would be equally distributed in the three scenarios presented using the different NCTs, and hence, the conclusion of the results are likely to remain the same. Finally, the use of different food Se content (i.e. using other FCTs for the underlying data), may lead to different estimates of apparent Se intakes and inadequacy. Nevertheless, this supports the hypothesis that it is essential to transparently report the food composition data underlying the estimates, and it is likely that small area Se composition would be more representative that Se values from different geographies. Thus, the main conclusion of the study would still be that localised Se concentration values are essential when estimating the prevalence of Se inadequacy in Malawi.

Conclusions

The use of small area Se concentration in maize, in context where food systems are highly localised and dietary diversity is low, would help to identify vulnerable populations and to target the intervention to areas in the country that are most in need. More efforts should be made to collect georeferenced Se composition data, and other mineral micronutrients, especially in staple crops and foods. Furthermore, estimates of dietary Se intakes and inadequacy risks using imputed values should be avoided or interpreted with appropriate caution. Relatedly, dietary Se estimates should be accompanied by sufficient metadata and methodological transparency, such that researchers and others can identify the origin and relevance of the Se composition values used.

References

Abdu, A. O., Kumssa, D. B., Joy, E. J. M., Groote, H. D., Lark, R. M., Broadley, M. R., & Gashu, D. (2022). Estimates of Dietary Mineral Micronutrient Supply from Staple Cereals in Ethiopia at a District Level. *Nutrients*, *14*(17), 3469. https://doi.org/10.3390/nu14173469

Allen, L. H., Carriquiry, A. L., & Murphy, S. P. (2020). Perspective: Proposed Harmonized Nutrient Reference Values for Populations. *Advances in Nutrition*, *11*(3), 469–483. https://doi.org/10.1093/advances/nmz096

Broadley, M. R., Chilimba, A. D. C., Joy, E. J. M., Young, S. D., Black, C. R., Ander, L. E., Watts, M. J., Hurst, R., Fairweather-Tait, S. J., White, P. J., & Gibson, R. S. (2012). Dietary requirements for magnesium, but not calcium, are likely to be met in Malawi based on national food supply data. *International Journal for Vitamin and Nutrition Research*, *82*(3), 192–199. Africa-Wide Information.

Carriquiry, A. L. (1999). Assessing the prevalence of nutrient inadequacy. *Public Health Nutrition*, *2*(1), 23–33. https://doi.org/10.1017/s1368980099000038

Charrondière, U. R., Stadlmayr, B., Grande, F., Vincent, A., Oseredczuk, M., Sivakumaran, S., Puwastien, P., Judprasong, K., Haytowitz, D., & Gnagnarella, P. (2023). *FAO/INFOODS Evaluation framework to assess the quality of published food composition tables and databases: User guide*. FAO. https://doi.org/10.4060/cc5371en

Chilimba, A. D. C., Young, S. D., Black, C. R., Rogerson, K. B., Ander, E. L., Watts, M. J., Lammel, J., & Broadley, M. R. (2011). Maize grain and soil surveys reveal suboptimal dietary selenium intake is widespread in Malawi. *Scientific Reports*, *1*(1), 72. https://doi.org/10.1038/srep00072

Chiutsi-Phiri, G., Kalimbira, A. A., Banda, L., Nalivata, P. C., Sanuka, M., Kalumikiza, Z., Matandika, L., Mfutso-Bengo, J., Allen, E., Ferguson, E., Sturgess, J., Broadley, M. R., Langley-Evans, S., Millar, K., Gashu, D., & Joy, E. J. M. (2021). Preparing for a community-based agriculture-to-nutrition trial in rural Malawi: Formative research to assess feasibility and inform design and implementation decisions. *Pilot and Feasibility Studies*, *7*(1), 141. https://doi.org/10.1186/s40814-021-00877-1

Combs, G. F. (2015). Biomarkers of Selenium Status. *Nutrients*, *7*(4), Article 4. https://doi.org/10.3390/nu7042209

Ecker, O., & Qaim, M. (2011). Analyzing Nutritional Impacts of Policies: An Empirical Study for Malawi. *World Development*, *39*(3), 412–428. https://doi.org/10.1016/j.worlddev.2010.08.002

Ekpa, O., Palacios-Rojas, N., Kruseman, G., Fogliano, V., & Linnemann, A. R. (2019). Sub-Saharan African Maize-Based Foods—Processing Practices, Challenges and Opportunities. *Food Reviews International*, *35*(7), 609–639. https://doi.org/10.1080/87559129.2019.1588290

FAO (1953). *Food composition tables for international use*. https://www.fao.org/4/X5557E/x5557e00.htm

FAO, & Government of Kenya. (2018). *Kenya Food Composition Tables*. http://www.fao.org/3/I9120EN/i9120en.pdf

FAO & The World Bank. (2018). *Food Data Collection in Household Consumption and Expenditure Surveys: Guidelines for Low- and Middle-Income Countries*. The World Bank Group.

FAO/INFOODS. (2012). *FAO / INFOODS Guidelines for Food Matching Version 1.2*. https://www.fao.org/3/ap805e/ap805e.pdf

Gashu, D., Nalivata, P. C., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A. E., Mossa, A. W., Munthali, M., … Broadley, M. R. (2021). The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*, *594*(7861), Article 7861. https://doi.org/10.1038/s41586-021-03559-3

Gibson, R. S., & Ferguson, E. (2008). *An interactive 24-hour recall for assessing the adequacy of iron and zinc intakes in developing countries*. 160.

Gwirtz, J. A., & Garcia-Casal, M. N. (2014). Processing maize flour and corn meal food products. *Annals of the New York Academy of Sciences*, *1312*(1), 66–75. https://doi.org/10.1111/nyas.12299

Hurst, R., Siyame, E. W. P., Young, S. D., Chilimba, A. D. C., Joy, E. J. M., Black, C. R., Ander, E. L., Watts, M. J., Chilima, B., Gondwe, J., Kang'ombe, D., Stein, A. J., Fairweather-Tait, S. J., Gibson, R. S., Kalimbira, A. A., & Broadley, M. R. (2013). Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi. *Scientific Reports*, *3*, 1425–1425. Africa-Wide Information. https://doi.org/10.1038/srep01425

Joy, E. J. M., Broadley, M.R., Young, S.D., Black, C.R., Chilimba, A.D.C., Ander, E.L., Barlow, T.S., Watts, M.J. (2015). Soil type influences crop mineral composition in Malawi. Science of the Total Environment 505, 587–595. https://doi.org/10.1016/j.scitotenv.2014.10.038

Joy, E. J. M., Kumssa, D.B., Broadley, M.R., Watts, M.J., Young, S.D., Chilimba, A.D.C., Ander, E.L. (2015). Dietary mineral supplies in Malawi: spatial and socioeconomic assessment. BMC Nutrition 1, 42. https://doi.org/10.1186/s40795-015-0036-4

Kumssa, D. B., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Chimungu, J. G., Davis, K., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A., Muleya, P., Munthali, M., … Nalivata, P. C. (2022). Cereal grain mineral micronutrient and soil chemistry data from GeoNutrition surveys in ethiopia and malawi. *Scientific Data*, *9*(1), Article 1. https://doi.org/10.1038/s41597-022-01500-5

Lephole, M., Khaketla, M., & Monoto, M. (2006). *Lesotho Food Composition Table*. Maseru: Department of Agricultural Research.

Ligowe, I. S., Phiri, F. P., Ander, E. L., Bailey, E. H., Chilimba, A. D. C., Gashu, D., Joy, E. J. M., Lark, R. M., Kabambe, V., Kalimbira, A. A., Kumssa, D. B., Nalivata, P. C., Young, S. D., & Broadley, M. R. (2020). Selenium deficiency risks in sub-Saharan African food systems and their geospatial linkages. *Proceedings of the Nutrition Society*, *79*(4), 457–467. https://doi.org/10.1017/S0029665120006904

Lividini, K., & Masters, W. A. (2022). Tracing global flows of bioactive compounds from farm to fork in nutrient balance sheets can help guide intervention towards healthier food supplies. *Nature Food*, *3*(9), 703–715. https://doi.org/10.1038/s43016-022-00585-w

Mlambo, L., Segovia de la Revilla, L., Codd, T., Osman, G., Tang, K., Joy, E. J. M., & Ander, L. E. (2024). *R for Household Consumption and Expenditure Surveys*. https://www.maps.africa/r4hces/

Mlotha, V., Mwangwela, A. M., Kasapila, W., Siyame, E. W. P., & Masamba, K. (2016). Glycemic responses to maize flour stiff porridges prepared using local recipes in Malawi. *Food Science & Nutrition*, *4*(2), 322–328. https://doi.org/10.1002/fsn3.293

Moltedo, A., Alvarez-Sanchez, C., Troubat, N., & Cafiero, C. (2022). *Estimating the prevalence of nutrient inadequacy from household consumption and expenditure surveys*. https://doi.org/10.4060/cc2780en

National Statistical Office. (2017). *Fourth Integrated Household Survey (IHS4) 2016-2017. Basic Information Document*.

Phiri, F. P., Ander, E. L., Bailey, E. H., Chilima, B., Chilimba, A. D. C., Gondwe, J., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Phuka, J. C., Salter, A., Suchdev, P. S., Watts, M. J., Young, S. D., & Broadley, M. R. (2019). The risk of selenium deficiency in Malawi is large and varies over multiple spatial scales. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-43013-z

Public Health England. (2021). McCance and Widdowson's composition of foods integrated dataset.

Segovia de la Revilla, L., Codd, T., Joy, E. J. M., Mlambo, L., Grande, F., Rittenschober, D., Moltedo, A., Holmes, B. A., & Ander, E. L. (2025). An open science framework and tools to create reproducible food composition data for use in nutrition. *Journal of Food Composition and Analysis*, *137*, 106894. https://doi.org/10.1016/j.jfca.2024.106894

Segovia de la Revilla, L., Ferguson, E. L., Dooley, C., Osman, G., Ander, E. L., & Joy, E. J. M. (2023). The availability and geographic location of open-source food composition data used to estimate micronutrient intakes in sub-Saharan Africa: A scoping review. *Journal of Food Composition and Analysis*, *120*, 105322. https://doi.org/10.1016/j.jfca.2023.105322

Stevens, G. A., Beal, T., Mbuya, M. N. N., Luo, H., Neufeld, L. M., Addo, O. Y., Adu-Afarwuah, S., Alayón, S., Bhutta, Z., Brown, K. H., Jefferds, M. E., Engle-Stone, R., Fawzi, W., Hess, S. Y., Johnston, R., Katz, J., Krasevec, J., McDonald, C. M., Mei, Z., … Young, M. F. (2022). Micronutrient deficiencies among preschool-aged children and women of reproductive age worldwide: A pooled analysis of individual-level data from population-representative surveys. *The Lancet Global Health*, *10*(11), e1590–e1599. https://doi.org/10.1016/S2214-109X(22)00367-9

Tang, K., Adams, K. P., Ferguson, E. L., Woldt, M., Kalimbira, A. A., Likoswe, B., Yourkavitch, J., Chrisinger, B., Pedersen, S., Segovia De La Revilla, L., Dary, O., Ander, E. L., & Joy, E. J. M. (2021). Modeling food fortification contributions to micronutrient requirements in Malawi using Household Consumption and Expenditure Surveys. *Annals of the New York Academy of Sciences*, *n/a*(n/a). https://doi.org/10.1111/nyas.14697

Tang, K., Adams, K. P., Ferguson, E. L., Woldt, M., Yourkavitch, J., Pedersen, S., Broadley, M. R., Dary, O., Ander, E. L., & Joy, E. J. M. (2022). Systematic review of metrics used to characterize dietary nutrient supply from household consumption and expenditure surveys. *Public Health Nutrition*, 1–32. https://doi.org/10.1017/S1368980022000118

United States Department of Agriculture (USDA), 2019. Agricultural Research Service, Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Legacy. Version Current: April 2019. Washington, DC, Agricultural Research Service, Nutrient Data Laboratory. https://fdc.nal.usda.gov/ download-datasets.html.

van Graan, A., Joelaine, C., Malory, J., Sitilitha, M., Agnes, M., Felix P., P., Lynne M., A., Shibani, G., & Elizabeth, M.-C. (2019). *Malawian Food Composition Tables*.

Watts, M. J., Joy, E. J. M., Young, S. D., Broadley, M. R., Chilimba, A. D. C., Gibson, R. S., Siyame, E. W. P., Kalimbira, A. A., Chilima, B., & Ander, E. L. (2015). Iodine source apportionment in the Malawian diet. *Scientific Reports*, *5*(1), 15251. https://doi.org/10.1038/srep15251

Watts, M. J., Middleton, D. R. S., Marriott, A. L., Humphrey, O. S., Hamilton, E. M., Gardner, A., Smith, M., McCormack, V. A., Menya, D., Munishi, M. O., Mmbaga, B. T., & Osano, O. (2019). Source apportionment of micronutrients in the diets of Kilimanjaro,Tanzania and Counties of Western Kenya. *Scientific Reports*, *9*(1), 14447–14447. https://doi.org/10.1038/s41598-019-51075-2

Weisell, R., & Dop, M. C. (2012). The adult male equivalent concept and its application to Household Consumption and Expenditures Surveys (HCES). *Food and Nutrition Bulletin*, *33*(3 Suppl), S157-162. https://doi.org/10.1177/15648265120333S203

Wheeler, E. F. (1991). Intra-household food and nutrient allocation. *Nutrition Research Reviews*. https://doi.org/10.1079/NRR19910008

World Health Organization & United Nations Children's Fund (UNICEF). (2022). *National Nutrition Information Systems; Technical Note: Nutrition Data Value Chain*. https://iris.who.int/bitstream/handle/10665/375716/9789240071919-eng.pdf?sequence=1

Chapter 6: Supplementary Materials

Supplementary Table 1. The baseline Nutrient Conversion Table (NCT). It provides the food matches between the food list of the Malawi 2016-17 Integrated Household Survey, Wave 4 (code, item) and various Food Composition Tables (FCTs) food items (fdc_id, food_desc), the energy, water and Se values used.

Supplementary Table 2. The list of food items food list of the Malawi 2016-17 Integrated Household Survey, Wave 4 (code, item) and their corresponding food groups (FoodName_1).

Supplementary Table 3. The percentage of households consuming each food group, their median (IQR) consumption (in g/AFE/day), the apparent energy (kcal/AFE/ day) and Se intakes (µg/AFE/day) by each NCT.

Supplementary Table 4. The percentage of population at risk of inadequacy based on the three apparent intake of Se scenarios: Baseline NCT, National (maize Se) NCT, and small area (maize Se) NCT.

Supplementary Figure 1. Comparison of the apparent median Se intakes (in µg/AFE/day) estimated using the baseline NCT (yellow), national maize NCT (dark blue) and small area maize NCT (light blue) per each district and region in Malawi.

code	item	source_f	fdc_id	ref_fctitem	WATERg	ENERCkcal	SEmcg	comments
101	Maize ufa mgaiwa (normal flour)	KE18	1022	Maize, whole, flour, raw	13.6	345	9.00	
102	Maize ufa refined (fine flour)	KE18	1022	Maize, whole, flour, raw	13.6	345	7.65	SEmcg value from KE18(1022) and applied extraction rates from (FAO,1953) Meal, fine, bolted and degermited 85%
103	Maize ufa madeya (bran flour)	KE18	1022	Maize, whole, flour, raw	13.6	345	9.00	
104	Maize grain (not as ufa)	MW19	MW01_0037	Maize, grain, white, raw, (Chimanga)	10.9	370	1.61	
105	Green maize	KE18	1023	Green Maize, white, whole, grain, fresh, raw	55.1	157	5.00	
106	Rice	KE18	1034	Rice, white, milled, polished grain, dry, raw	12.2	353	1.00	
107	Finger millet (mawere)	MW19	MW01_0016	Finger millet, grain, raw, (Mawere)	8.7	378	3.02	
108	Sorghum (mapira)	MW19	MW01_0060	Sorghum, wholegrain, raw, Sorghum bicolor, (Mapira)	12.4	358	19.80	
109	Pearl millet (mchewere)	KE18	1025	Millet, bulrush, grain, dry, raw	11.0	354	25.00	
110	Wheat flour	KE18	1044	Wheat, whole, flour, raw	12.5	328	48.00	
111	Bread	KE18; MW19	1007; MW01_0003	Bread, White; Bread, wheat, brown, homemade	24.1	305	5.00	
112	Buns, scones	KE18; KE18	1036; 1011	Scone, plain, commercial; Buns, Currant	25.0	341	5.50	
113	Biscuits	KE18	1004	Biscuit, Sweet	4.3	460	10.00	
114	Spaghetti, macaroni, pasta	KE18	1031	Pasta, spaghetti, dry, raw-imported	10.6	354	12.00	
115	Breakfast cereal	KE18	1009	Breakfast cereal, flakes of corn	3.5	373	9.00	
116	Infant feeding cereals	MW19;MW19;MW19;MW19;MW19;MW19;MW19;MW19	MW07_0001;MW07_0002;MW07_0003;MW07_0004;MW07_0005;MW07_0006;MW07_0007;MW07_0008	baby cereal, containing milk, 12 months, strawberry flavour, dry;baby cereal, containing milk, 6 months, dry;baby cereal, containing milk, 7 months, regular flavour, dry;baby cereal, containing milk, 9 months, mixed fruit flavour, dry;baby cereal, maize, 6 months, dry;baby cereal, maize, 9 months, strawberry and banana flavour, dry (nestum);baby cereal, mixed cereal, 7 months, dry;baby cereal, mixed grain, 7 months, dry	4.6	385	25.90	
201	Cassava tubers	MW19	MW01_0011	Cassava, tuber, raw, (Chinangwa chachiwisi)	59.7	160	0.25	
202	Cassava flour	MW19	MW01_0017	flour, cassava, raw, (ufa wa kondowole)	14.0	348	2.00	
203	White sweet potato	MW19	MW01_0065	Sweet potato, white-fleshed, raw, Ipomoea batatas, (Mbatata zoyera mkati)	77.3	89	0.04	
204	Orange sweet potato	KE18	2014	Sweetpotato, orange, biofortified, raw	78.7	77	1.00	

code	item	source_f	fdc_id	ref_fctitem	WATERg	ENERCkcal	SEmcg	comments
205	Irish potato	KE18	2009	Potato, Irish (English), white variety, peeled, raw	72.1	105	1.00	
206	Potato crisps	KE18	15110	Potato Chips	32.1	326	1.00	
207	Plantain, cooking banana	KE18	2004	Banana, plantain, green, raw	73.6	97	1.00	
208	Cocoyam (masimbi)	KE18	2016	Taro, peeled, raw	73.0	97	1.00	
301	Bean, white	KE18	3005	Beans, lima, dry, raw	13.6	299	6.00	
302	Bean, brown	MW19	MW02_0004	Bean, kidney, dry, Phaseolus vulgaris, (Nyemba)	11.0	343	10.38	
303	Pigeonpea (nandolo)	KE18	3021	Pigeon peas, dry, raw	12.1	303	15.00	
305	Groundnut flour	MW19	MW02_0010	flour, groundnut, arachis hypogaea, (nsinjilo)	8.0	583	3.05	
306	Soyabean flour	MW19	MW02_0012	Flour, roasted soybean, (Ufa wa soya)	9.1	458	7.40	
307	Ground bean (nzama)	KE18	3001	Beans, broad, dry, raw	10.9	305	8.00	
308	Cowpea (khobwe)	MW19	MW02_0007	Cowpea, dry, Vigna unguiculata, (Khobwe/Nseula)	11.1	349	5.80	
309	Macademia nuts	KE18	10010	Nut, macadamia, raw, unsalted	7.7	696	10.00	
311	Groundnut (shelled)	MW19	MW02_0014	Groundnut, dry, Arachis hypogae, (Mtedza wouma)	6.5	597	3.05	
312	Groundnut (unshelled)	MW19	MW02_0014	Groundnut, dry, Arachis hypogae, (Mtedza wouma)	6.5	597	3.05	
313	Groundnut fresh (unshelled)	MW19	MW02_0014	Groundnut, dry, Arachis hypogae, (Mtedza wouma)	6.5	597	3.05	
314	Soya	MW19	MW02_0019	soybean, dry, (soya)	8.5	432	5.25	
401	Onion	MW19	MW04_0031	Onion, raw, (Anyezi)	89.1	43	0.32	
402	Cabbage	MW19	MW04_0004	Cabbage, raw, (Kabichi)	92.2	32	0.70	
403	Tanaposi/Rape	MW19	MW04_0020	Leaves, rape, raw, (Lepu)	89.7	37	0.84	
404	Nkhwani	MW19	MW04_0019	Leaves, pumpkin, raw, (Nkhwani)	92.9	27	0.56	
405	Chinese cabbage	KE18	4005	Cabbage, leaf head, Chinese, raw	97.0	10	1.00	
406	Other cultivated green leafy vegetables	KE18; MW19	4018; MW04_0011	Jute mallow, picked leaves, raw; Leaves, amaranth, raw, (Bonongwe)	85.1	43	0.97	
407	Gathered wild green leaves	KE18; KE18; MW19	4029; 4018; MW04_0012	Spider plant, leaves, raw; Jute mallow, picked leaves, raw; Leaves, black jack, raw, (Chisoso)	85.0	43	1.00	
408	Tomato	MW19	MW04_0036	FALSE	94.5	26	0.43	
409	Cucumber	KE18	4016	Cucumber, green, unpeeled, raw	96.2	11	1.00	
410	Pumpkin	KE18	4027	Pumpkin, flesh, yellow w/o seeds, raw	90.3	32	1.00	
411	Okra / Therere	KE18	4026	Okra, fresh, raw	90.0	32	1.00	
412	Tinned vegetables (Specify)	KE18	4025	Mushroom, raw, canned in brine (drained)	93.0	18	16.00	
413	Mushroom	MW19	MW04_0025	Mushroom, oyster, cultivated, raw, Pleurotus florida,(Bowa wolimidwa)	91.7	30	1.99	
501	Eggs	KE18	7011	Egg, chicken, whole, rawÃ,Â Ã,Â Ã,Â	75.9	134	23.00	

code	item	source_f	fdc_id	ref_fctitem	WATERg	ENERCkcal	SEmcg	comments
504	Beef	KE18; KE18; KE18	7001; 7002; 7004	Beef, high fat, w/o bones, rawÃ,Â Ã,Â Ã,Â Ã,Â Ã,Â ; Beef, lean, raw; Beef, medium fat, w/o bones, raw	70.3	151	15.00	
505	Goat	KE18	7016	Goat, medium fat, raw	68.2	166	26.00	
506	Pork	KE18	7020	Pork, meat, raw (unspecified part)	57.2	286	28.00	
507	Mutton	KE18	7019	Lamb, raw (unspecified part)	60.0	269	20.00	
508	Chicken	MW19	MW03_0011	chicken, meat with skin, free range, local, raw, (nkhuku ya chikuda)	73.1	129	11.84	
509	Other poultry - guinea fowl, doves, etc.	KE18; KE18	7026; 7017	Quail, flesh & skin, raw; Guinea fowl, meat, with skin	72.5	144	20.00	
510	Small animal Ã¯Â¿Â½ rabbit, mice, etc.	KE18	7021	Rabbit meat, raw	67.0	151	25.00	
511	Termites, other insects (eg Ngumbi, caterpillar)	KE18	14003	Termite, Dry, Raw	7.0	557	4.00	
512	Tinned meat or fish	UK21	16-439	Sardines, canned in sunflower oil, drained	58.6	220	49.00	
514	Fish Soup/Sauce	MW19	MW03_0019	Fish powder stew, (Supu ya ufa wansomba)	91.1	44	6.01	
522	Chicken/Pieces	KE18	7009	Chicken, unespecified part, w/o bone, meat&skin, raw	63.6	207	16.00	
601	Mango	MW19	MW05_0016	Mango,ripe, (Mango)	83.5	66	0.77	
602	Banana	KE18	5004	Banana, cavendish, raw	74.4	95	0.00	
603	Citrus Ã¯Â¿Â½ naartje, orange, etc.	US19; KE18; KE18; KE18	9218; 5023; 5015; 5017	Tangerines, (mandarin oranges), raw; Orange, pulp, raw; Lemon, pulp, raw; Lime, pulp, raw	87.4	42	0.00	
604	Pineapple	KE18	5030	Pineapple, raw	88.6	41	1.00	
605	Papaya	MW19	MW05_0019	Pawpaw, fresh, (Papaya)	88.1	48	1.29	
606	Guava	KE18	5011	Guava, pink-fleshed, raw	80.7	48	0.00	
607	Avocado	KE18	5003	Avocado, ripe, raw	74.0	185	0.00	
608	Wild fruit (masau, malambe, etc.)	MW19;MW19;MW19;MW19;MW19;MW19;MW19	MW05_0013;MW05_0005;MW05_0006;MW05_0009;MW05_0015;MW05_0017;MW05_0024	jujube, ziziphus mauritiana, (masau);baobab, raw, adansonia digitata,(malambe);custard apple, wild, annona senegalensis, (mpoza);jakjak fruit, azanza garckeana, (matowo);loquats, wild, uapaca kirkiana,(masuku a mtchire);medlar, african, raw, vangueria infausta, (lokwati);tamarind, fruit, raw, tamarindus indica, (bwemba)	57.7	167	4.03	
609	Apple	MW19	MW05_0001	apple, average, raw, (apozi)	83.9	63	1.07	
701	Fresh milk	KE18	6022	Milk, cow, whole, fresh, raw	86.3	70	3.00	
702	Powdered milk	KE18	6018	Milk, cow, powder, whole	3.2	495	25.00	
703	Margarine - Blue band	KE18	9005	Margarine,60% fat	37.7	541	0.00	
704	Butter	KE18	6001	Butter (cow milk), no added salt (~80-84% fat)	17.2	735	1.00	
705	Chambiko - soured milk	KE18	6021	Milk, cow, whole, fermented (Lala - mursik)	89.1	58	0.00	

code	item	source_f	fdc_id	ref_fctitem	WATERg	ENERCkcal	SEmcg	comments
706	Yoghurt	KE18	6026	Yoghurt, cow milk, whole, plain	82.0	85	2.00	
707	Cheese	KE18	6006	Cheese, cottage, Milk, Cow, Sour	73.6	122	10.00	
801	Sugar	KE18	11003	Sugar, white, granulated or lump	0.0	400	1.00	
802	Sugar Cane	MW19	MW08_0007	sugarcane, raw, saccharum officinarum, (mzimbe)	90.0	40	0.00	
803	Cooking oil	KE18; KE18; KE18	9013; 9012; 9009	Sun flower oil; Soya bean oil; Peanut oil	0.0	900	0.00	
810	Salt	KE18	13027	salt, iodized	0.2	0	2.00	
811	Spices	KE18	13028	Spice, mixed or all spice	8.5	360	3.00	
812	Yeast, baking powder, bicarbonate of soda	US19; US19	43406; 18369	Yeast extract spread; Leavening agents, baking powder, double-acting, sodium aluminum sulfate	23.0	119	13.90	
813	Tomato sauce (bottle)	KE18	13031	Tomato, sauce (Ketchup)	67.5	115	0.50	
814	Hot sauce (Nali, etc.)	US19	6961	Sauce, peppers, hot, chili, mature red, canned	94.1	21	0.20	
815	Jam, jelly	US19	19300	Jellies	29.8	266	0.40	
816	Sweets, candy, chocolates	US19	19162	Candies, WHATCHAMACALLIT Candy Bar	3.0	494	0.70	
817	Honey	KE18	11001	honey, raw	19.2	322	0.78	
820	Maize - boiled or roasted (vendor)	KE18	1051	Green maize, white, whole, grain, fresh, boiled, drained (without salt)	58.0	146	5.00	
821	Chips (vendor)	KE18	15110	Potato Chips	32.1	326	1.00	
822	Cassava - boiled (vendor)	MW19	MW01_0010	cassava, tuber, boiled, (chinangwa chophika)	63.0	146	0.94	
823	Eggs - boiled (vendor)	KE18	7027	Egg, chicken, whole, boiled (without salt)	75.9	134	21.00	
824	Chicken (vendor)	KE18; MW19	15073; MW03_0010	Ingokho (Fried Chicken); Chicken stew, (Nkhuku yokazingira)	39.7	345	20.64	
825	Meat (vendor)	KE18; KE18	15073; 7014	Ingokho (Fried Chicken); Goat, lean, raw	40.7	326	30.50	
826	Fish (vendor)	KE18	8029	Mudfish (kamongo), fillet, grilled (without salt and fat)	71.8	133	49.00	
827	Mandazi, doughnut (vendor)	KE18	15003	Roti (Indian Chapati)	18.8	429	0.00	
828	Samosa (vendor)	KE18; KE18	15026; 15025	Vegetable Samosa (Sambusa ya Mboga); Meat Samosa (Sambusa ya Nyama)	22.1	384	7.00	
829	Meal eaten at restaurant (vendor)	MW19; MW19	MW03_0010; MW01_0031	Chicken stew, (Nkhuku yokazingira); Maize thick porridge, degermed-dehulled flour, (Nsima ya galamilu)	77.6	115	1.15	
831	Boiled sweet potatoes	MW19	MW01_0066	sweet potato, white-fleshed, without skin, boiled, ipomoea batatas, (mbatata yoyera mkati yowilitsa)	76.0	97	0.00	
832	Rosted sweet potatoes	MW19	MW01_0065	Sweet potato, white-fleshed, raw, Ipomoea batatas, (Mbatata zoyera mkati)	77.3	89	1.00	Semcg adjusted for retention factors based on KE18.

code	item	source_f	fdc_id	ref_fctitem	WATERg	ENERCkcal	SEmcg	comments
833	Boiled groundnuts	US19	16088	Peanuts, all types, cooked, boiled, with salt	41.8	318	4.40	
834	Rosted groundnuts	US19	16090	Peanuts, all types, dry-roasted, with salt	1.8	587	9.30	
835	popcorn	US19	167959	Snacks, popcorn, air-popped	3.3	387	0.00	
836	Zikondamoyo/Nkate	US19	18019	Bread, banana, prepared from recipe, made with margarine	29.2	326	12.10	
838	Cassava - roasted (vendor)	MW19	MW01_0010	cassava, tuber, boiled, (chinangwa chophika)	63.0	146	0.95	
901	Tea	KE18	12005	tea, chai, instant dry powder	5.1	299	15.00	
902	Coffee	KE18	12003	coffee, instant, dry powder or granules	3.1	311	9.00	
903	Cocoa, millo	KE18	12004	drinking chocolate, powder	1.0	541	3.00	
904	Squash (Sobo drink concentrate)	UK21	17-738	Fruit juice drink/squash, undiluted	91.0	34	0.00	
905	Fruit juice	MW19	MW05_0012	Juice, orange, homemade, (Juwisi wa malalanje)	77.8	89	0.01	
906	Freezes (flavoured ice)	US19	19283	Frozen novelties, ice type, pop	80.5	79	0.20	
907	Soft drinks (Coca-cola, Fanta, Sprite, etc.)	US19	175113	Carbonated beverage, low calorie, other than cola or pepper, with sodium saccharin, without caffeine	99.8	0	0.00	
908	Chibuku (commercial traditional-style beer)	LSOFCT	140001	beer, 5% alcohol	92.0	16	0.60	
909	Bottled water	MW19	MW08_0008	Water, (Madzi)	100.0	0	0.02	
910	Maheu	UK21	17-749	Lager, standard	93.0	24	0.00	
911	Bottled / canned beer (Carlsberg, etc.)	UK21	17-749	Lager, standard	93.0	24	0.00	
912	Thobwa	UK21	17-749	Lager, standard	93.0	24	0.00	
913	Traditional beer (masese)	UK21	17-749	Lager, standard	93.0	24	0.00	
914	Wine or commercial liquor	KE18	12008	Wine, White, Dry	88.7	76	1.00	
915	Locally brewed liquor (kachasu)	US19	14037	Alcoholic beverage, distilled, all (gin, rum, vodka, whiskey) 80 proof	66.6	231	0.00	
5021	Sun Dried fish (Large Variety)	MW19;MW19	MW03_0031;MW03_0020	fish, tilapia, whole, dried, oreochromis shiranus, (chambo chouma);fish, catfish, dry, clarius gariepinus, (mlamba wouma)	11.5	375	109.05	SEmcg value from UF16(91005) water-adjusted.
5022	Sun Dried fish (Medium Variety)	MW19;MW19	MW03_0031;MW03_0020	fish, tilapia, whole, dried, oreochromis shiranus, (chambo chouma);fish, catfish, dry, clarius gariepinus, (mlamba wouma)	11.5	375	109.05	SEmcg value from UF16(91005) water-adjusted.
5023	Sun Dried fish (Small Variety)	MW19	MW03_0041	{(Fish, whole, salted, sun dried, Rhamphochromis esox, (Mcheni wadzuwa)}*	21.5	346	96.75	SEmcg value from KE18(8004) water-adjusted.
5031	Fresh fish (Large Variety)	KE18	8010	Nile tilapia, fillet, w/o skin and bones, raw	77.6	92	17.00	
5032	Fresh fish (Medium Variety)	KE18	8010	Nile tilapia, fillet, w/o skin and bones, raw	77.6	92	17.00	
5033	Fresh fish (Small Variety)	KE18	8004	Herring stock, raw	70.8	114	36.00	
5121	Smoked fish (Large Variety)	MW19;MW19	MW03_0047;MW03_0023	fish, tilapia, whole, smoked, sun dried, oreochromis shiranus, (chambo chowamba);fish, catfish, smoked, clarias gariepinus, (mlamba wowamba)	11.5	394	148.15	SEmcg value from UF16(91004) water-adjusted.
5122	Smoked fish (Medium Variety)	MW19;MW19	MW03_0047;MW03_0023	fish, tilapia, whole, smoked, sun dried, oreochromis shiranus, (chambo chowamba);fish, catfish, smoked, clarias gariepinus, (mlamba wowamba)	11.5	394	148.15	SEmcg value from UF16(91004) water-adjusted.
5123	Smoked fish (Small Variety)	MW19	MW03_0039	Fish, whole, par-boiled, sun dried, Engraulicypris sardella, (Usipa ofutsa)	54.5	194	56.07	SEmcg value from KE18(8004) water-adjusted.

code	item	FoodName_1
101	Maize ufa mgaiwa (normal flour)	maize and products (including white maize)
102	Maize ufa refined (fine flour)	maize and products (including white maize)
103	Maize ufa madeya (bran flour)	maize and products (including white maize)
104	Maize grain (not as ufa)	maize and products (including white maize)
105	Green maize	maize and products (including white maize)
106	Rice	rice and products
107	Finger millet (mawere)	millet and products
108	Sorghum (mapira)	sorghum and products
109	Pearl millet (mchewere)	millet and products
110	Wheat flour	wheat and products
111	Bread	wheat and products
112	Buns, scones	wheat and products
113	Biscuits	wheat and products
114	Spaghetti, macaroni, pasta	wheat and products
115	Breakfast cereal	wheat and products
116	Infant feeding cereals	infant food and products
201	Cassava tubers	cassava and products
202	Cassava flour	cassava and products
203	White sweet potato	sweet potatoes and products
204	Orange sweet potato	sweet potatoes and products
205	Irish potato	potatoes and products
206	Potato crisps	potatoes and products
207	Plantain, cooking banana	plantains and products
208	Cocoyam (masimbi)	roots, other and products
301	Bean, white	beans and products
302	Bean, brown	beans and products
303	Pigeonpea (nandolo)	pulses, other and products
305	Groundnut flour	Groundnuts
306	Soyabean flour	soyabeans and products
307	Ground bean (nzama)	pulses, other and products
308	Cowpea (khobwe)	pulses, other and products
309	Macademia nuts	nuts and products
311	Groundnut (shelled)	Groundnuts
312	Groundnut (unshelled)	Groundnuts
313	Groundnut fresh (unshelled)	Groundnuts
314	Soya	soyabeans and products
401	Onion	onions and products
402	Cabbage	vegetables, other and products
403	Tanaposi/Rape	vegetables, other and products
404	Nkhwani	vegetables, other and products
405	Chinese cabbage	vegetables, other and products
406	Other cultivated green leafy vegetables	vegetables, other and products
407	Gathered wild green leaves	vegetables, other and products

code	item	FoodName_1
408	Tomato	tomatoes and products
409	Cucumber	vegetables, other and products
410	Pumpkin	vegetables, other and products
411	Okra / Therere	vegetables, other and products
412	Tinned vegetables (Specify)	vegetables, other and products
413	Mushroom	vegetables, other and products
501	Eggs	eggs and products
504	Beef	bovine meat and products
505	Goat	mutton & and goat meat and products
506	Pork	pigmeat and products
507	Mutton	mutton & and goat meat and products
508	Chicken	poultry meat and products
509	Other poultry - guinea fowl, doves, etc.	meat, other and products
510	Small animal Ã¯Â¿Â½ rabbit, mice, etc.	meat, other and products
511	Termites, other insects (eg Ngumbi, caterpillar)	meat, other and products
512	Tinned meat or fish	meat, other and products
514	Fish Soup/Sauce	fish and products
522	Chicken/Pieces	poultry meat and products
601	Mango	fruits, other and products
602	Banana	bananas and products
603	Citrus Ã¯Â¿Â½ naartje, orange, etc.	oranges, mandarines and products
604	Pineapple	pineapples and products
605	Papaya	fruits, other and products
606	Guava	fruits, other and products
607	Avocado	fruits, other and products
608	Wild fruit (masau, malambe, etc.)	fruits, other and products
609	Apple	apples and products
701	Fresh milk	milk - (excluding excluding butter butter) and products
702	Powdered milk	milk - (excluding excluding butter butter) and products
703	Margarine - Blue band	oilcrops oil, other and products
704	Butter	butter, ghee and products
705	Chambiko - soured milk	milk - (excluding excluding butter butter) and products
706	Yoghurt	milk - (excluding excluding butter butter) and products
707	Cheese	milk - (excluding excluding butter butter) and products
801	Sugar	sugar (raw equivalent) and products
802	Sugar Cane	sugar cane and products
803	Cooking oil	oils and products
810	Salt	spices, other and products
811	Spices	spices, other and products
812	Yeast, baking powder, bicarbonate of soda	miscellaneous and products
813	Tomato sauce (bottle)	tomatoes and products
814	Hot sauce (Nali, etc.)	miscellaneous and products
815	Jam, jelly	sugar (raw equivalent) and products

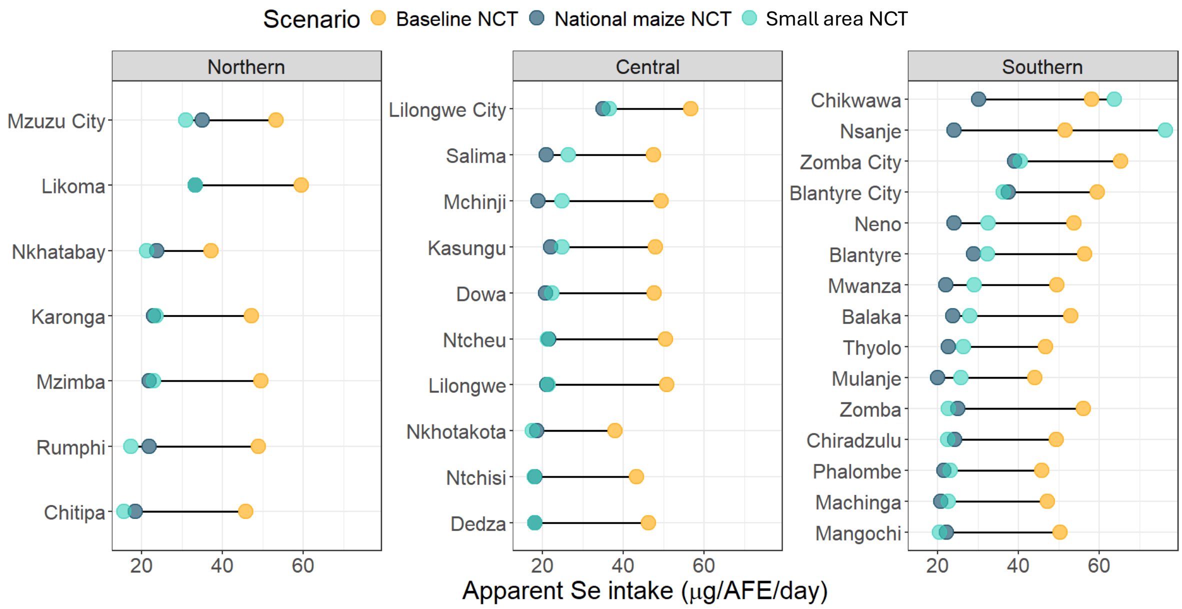
code	item	FoodName_1
816	Sweets, candy, chocolates	cocoa beans and products
817	Honey	honey and products
820	Maize - boiled or roasted (vendor)	maize and products (including white maize)
821	Chips (vendor)	potatoes and products
822	Cassava - boiled (vendor)	cassava and products
823	Eggs - boiled (vendor)	eggs and products
824	Chicken (vendor)	poultry meat and products
825	Meat (vendor)	miscellaneous and products
826	Fish (vendor)	fish and products
827	Mandazi, doughnut (vendor)	wheat and products
828	Samosa (vendor)	miscellaneous and products
829	Meal eaten at restaurant (vendor)	miscellaneous and products
831	Boiled sweet potatoes	sweet potatoes and products
832	Rosted sweet potatoes	sweet potatoes and products
833	Boiled groundnuts	Groundnuts
834	Rosted groundnuts	Groundnuts
835	popcorn	maize and products (including white maize)
836	Zikondamoyo/Nkate	miscellaneous and products
838	Cassava - roasted (vendor)	cassava and products
901	Tea	tea (including mate) and products
902	Coffee	coffee and products
903	Cocoa, millo	cocoa beans and products
904	Squash (Sobo drink concentrate)	sweeteners, other and products
905	Fruit juice	oranges, mandarines and products
906	Freezes (flavoured ice)	milk - (excluding excluding butter butter) and products
907	Soft drinks (Coca-cola, Fanta, Sprite, etc.)	sweeteners, other and products
908	Chibuku (commercial traditional-style beer)	beverages, fermented and products
909	Bottled water	miscellaneous and products
910	Maheu	beverages, fermented and products
911	Bottled / canned beer (Carlsberg, etc.)	beer and products
912	Thobwa	beverages, fermented and products
913	Traditional beer (masese)	beverages, fermented and products
914	Wine or commercial liquor	wine and products
915	Locally brewed liquor (kachasu)	beverages, alcoholic and products
5021	Sun Dried fish (Large Variety)	fish and products
5022	Sun Dried fish (Medium Variety)	fish and products
5023	Sun Dried fish (Small Variety)	fish and products
5031	Fresh fish (Large Variety)	fish and products
5032	Fresh fish (Medium Variety)	fish and products
5033	Fresh fish (Small Variety)	fish and products
5121	Smoked fish (Large Variety)	fish and products
5122	Smoked fish (Medium Variety)	fish and products
5123	Smoked fish (Small Variety)	fish and products

FoodName_1	perc	Q25_cons	median_cons	Q75_cons	mean_ener	sd_ener	enerc_per	median_ener	Q25_ener	Q75_ener
vegetables, other and products	98.32088	28.4284428	58.79352465	129.635604	36.6343356	50.04015	1.62296171	19.07493706	9.14183803	41.747122
spices, other and products	97.50944	5.00863558	8.803611738	15.625	1.17301691	9.921422	0.05196659	0	0	0
maize and products (including white maize)	97.28449	274.336283	391.0614525	564.516129	1494.25817	826.3673	66.1981104	1328.12532	936.961637	1888.16456
tomatoes and products	83.96401	18.2589614	30.65263158	51.6244264	10.5131767	9.827824	0.46575113	7.878118182	4.69183235	13.3536
oils and products	75.78533	4.8951049	11.95079086	22.9885058	154.314898	169.4559	6.8364054	107.5571178	44.0559441	206.896552
fish and products	73.64023	4.01904382	7.985071524	16.7344498	34.6353086	45.14338	1.5344015	21.00659643	11.1492763	39.5684177
sugar (raw equivalent) and products	55.59573	19.1374663	27.83204802	40.4624278	129.660135	87.2821	5.74415859	111.2177593	76.5498652	161.849711
wheat and products	52.3339	9.1603381	21.9541898	51.8029095	138.925972	186.2378	6.15465053	79.58811189	36.6700744	175.750366
beans and products	49.21668	16.188836	24.82517483	40.3458213	108.514494	96.57742	4.80737172	83.48928865	54.1038539	134.356033
onions and products	46.99928	3.84172662	6.625877166	11.420354	3.95218166	4.570397	0.17508819	2.816792901	1.63319482	4.85502089
fruits, other and products	38.18591	24.2608696	44.65408805	79.6261682	55.0385688	59.06057	2.4382997	35.92621224	18.5087177	68.4344538
Groundnuts	36.39431	3.76983803	8.497079395	20.6251813	118.892061	257.7424	5.26711509	49.58008511	21.776075	120.020525
sweet potatoes and products	32.47369	50.1322978	81.02564103	132.587413	92.326358	81.4776	4.09021049	69.30630648	42.8882348	113.308099
eggs and products	31.93541	7.36097228	10.89140271	16.7928155	18.8435741	15.21791	0.83480152	14.59447964	9.86370285	22.5023728
miscellaneous and products	31.24448	1.10803324	7.086614173	14.4	21.6799249	42.33669	0.96045656	10.2270364	1.20303286	23.8672316
pulses, other and products	31.19627	15.3248417	27.07930368	47.7821266	120.682255	125.5815	5.34642366	84.72703063	48.0304297	149.169633
cassava and products	27.49257	37.2058954	63.12138728	130.298208	277.830312	443.2472	12.3083427	102.8335194	60.1267261	254.915849
rice and products	24.96184	29.6513554	47.61904762	83.2139446	247.635844	265.4653	10.9706778	168.0952381	104.669285	293.745225
tea (including mate) and products	24.43962	0.37139385	0.76472071	1.62100457	4.92083002	8.826166	0.21800092	2.286514922	1.1104676	4.84680365
beverages, fermented and products	21.25813	46.2704649	107.8652291	219.715909	38.4408224	40.89057	1.70299206	24.76868327	10.7089118	50.8702679
milk - (excluding excluding butter butter) and products	21.17779	10.104304	26.33618931	58.8265576	56.6189164	76.63347	2.50831172	34.26573427	15.850326	70
bananas and products	19.4344	13.0952988	21.16465863	36.3364763	29.6358894	32.72174	1.31291896	20.1064257	12.4405339	34.5196525
sugar cane and products	18.2775	60.2678571	104.4642857	188.290051	59.7590363	55.92613	2.64742423	41.48485714	23.9335714	74.7737452
potatoes and products	18.00434	23.0769231	51.36842105	107.939189	109.097093	142.3208	4.83318179	68.80601504	35.5057804	126.267606
poultry meat and products	14.86302	27.5605624	40	61.1593879	68.4706016	57.6116	3.03336099	54.24638756	38.1095016	80.48125
mutton & and goat meat and products	14.54969	21.0526316	29.41176471	45.6026059	65.7853579	61.72379	2.91440025	48.92631579	34.9603311	76.1967213
oranges, mandarines and products	12.57331	17.5300172	33.58	60.9756098	31.717486	37.73827	1.40513713	19.08470907	7.70688525	40.7566058
soyabeans and products	11.98682	13.6178107	34.02906917	70.0544794	236.764546	255.4569	10.4890613	155.8531368	62.369573	320.849516
bovine meat and products	10.42822	24.3161232	32.99944506	48.951049	62.7031814	41.22127	2.77785473	49.82916204	36.717346	73.9160839
sweeteners, other and products	9.247208	25.3932929	45.11278195	78.125	12.9413389	14.50844	0.57332274	10.2	0	21.1618257
pigmeat and products	6.314775	17.0983016	24.82517483	43.2875075	126.928012	216.9428	5.62312105	71	48.9011426	123.802272
oilcrops oil, other and products	5.101631	6.22844828	10.34985423	16.2162162	75.3616906	98.20284	3.33864764	55.99271137	33.6959052	87.7297297
cocoa beans and products	4.989154	1.20786517	1.919642857	3.67170626	17.4378221	24.67907	0.77252438	9.483035714	5.98432304	18.7940447
sorghum and products	4.780268	39.4168684	108.3916084	213.888478	542.316489	538.7587	24.0255182	387.578042	140.943685	764.805308
plantains and products	4.241986	25.2493614	41.2486888	77.5873077	74.4923023	109.1875	3.3001323	40.01122813	24.4918805	75.2596885
meat, other and products	3.719772	2.79414032	4.980842912	10.4543249	43.6155213	86.10982	1.93223979	15.01347709	5.25493508	43.5362485
millet and products	2.353981	6.0952381	12.8	55.1181102	248.248465	476.1836	10.9978179	48.414208	22.7055394	204.777452
apples and products	2.01655	3.85714286	8.181818182	16.8711679	8.21436149	8.45117	0.36390981	5.146363636	2.42614286	10.6119646
beverages, alcoholic and products	1.896039	17.9580096	30.3937329	66.8314045	128.350045	164.778	5.68611943	70.20952301	41.4830021	154.380544
coffee and products	1.775528	0.5631068	1.492146597	3.95480226	11.3611357	19.22634	0.50331712	4.640575916	1.75126214	12.299435
wine and products	0.98819	17.3346402	35.34303534	63.8314176	47.0700298	66.54949	2.08528023	26.86070686	13.1743265	48.5118774
beer and products	0.763236	19.9152322	53.50553506	132.824726	27.0564712	40.27964	1.19864646	12.84132841	4.77965574	31.8779343
pineapples and products	0.755202	16.1146952	29.68841778	48.8388924	15.4409929	12.15448	0.68406154	12.17225129	6.60702505	20.0239459
roots, other and products	0.747168	81	153.4736842	248.882682	194.161965	177.3611	8.60169646	148.8694737	78.57	241.416201
butter, ghee and products	0.642725	8.31070372	12.90913358	20.1946386	141.986472	177.6317	6.29023572	94.88213184	61.0836723	148.430594
honey and products	0.634691	6.63003663	12	21.8944099	54.0800509	50.38816	2.39583577	38.64	21.348718	70.5
nuts and products	0.160681	5.55963683	10.35804473	21.1788691	127.139266	134.1054	5.63247992	72.09199134	38.6950723	147.404929

FoodName_1	median_Se	Q25_Se	Q75_Se	median_SeN	Q25_SeN	Q75_SeN	median_Se_ea	Q25_Se_ea	Q75_Se_ea
vegetables, other and products	0.49816199	0.23318	1.154305	0.498161985	0.2331795	1.1543055	0.498161985	0.233179545	1.154305459
spices, other and products	0.00916031	0.005098	0.016667	0.009160305	0.005098	0.0166667	0.009160305	0.005098039	0.016666667
maize and products (including white maize)	31.9370629	22.47946	45.84	7.023990474	4.4787717	10.527236	7.072744502	3.672658564	13.44616741
tomatoes and products	0.13183697	0.078532	0.222066	0.131836968	0.0785318	0.2220658	0.131836968	0.078531793	0.222065816
oils and products	0	0	0	0	0	0	0	0	0
fish and products	5.92469766	3.09375	11.22556	5.924697657	3.09375	11.225559	5.924697657	3.09375	11.22555905
sugar (raw equivalent) and products	0.2780441	0.191375	0.404041	0.278044104	0.1913747	0.4040406	0.278044104	0.191374663	0.404040551
wheat and products	0.83029795	0	2.561421	0.830297954	0	2.5614209	0.830297954	0	2.561420885
beans and products	2.4127907	1.504805	3.832454	2.412790698	1.5048048	3.8324538	2.412790698	1.50480478	3.832453826
onions and products	0.02144995	0.012437	0.036971	0.021449952	0.0124368	0.0369711	0.021449952	0.012436822	0.036971112
fruits, other and products	0.20262611	0	0.551571	0.202626111	0	0.5515708	0.202626111	0	0.551570793
Groundnuts	0.26333669	0.116702	0.639941	0.26333669	0.1167016	0.6399411	0.26333669	0.116701623	0.639941135
sweet potatoes and products	0.05338029	0.025989	0.446156	0.053380293	0.0259889	0.446156	0.053380293	0.025988947	0.446156019
eggs and products	2.50125	1.690532	3.8623	2.50125	1.6905319	3.8622998	2.50125	1.690531943	3.862299795
miscellaneous and products	0.83959732	0.130542	1.731488	0.839597315	0.1305423	1.7314879	0.839597315	0.130542328	1.731487889
pulses, other and products	3.49417637	1.704527	6.167406	3.494176373	1.7045267	6.1674056	3.494176373	1.704526686	6.167405618
cassava and products	0.16895628	0.095774	0.52897	0.168956284	0.0957736	0.5289698	0.168956284	0.095773578	0.528969776
rice and products	0.47619048	0.296514	0.832139	0.476190476	0.2965136	0.8321394	0.476190476	0.296513554	0.832139446
tea (including mate) and products	0.11470811	0.055709	0.243151	0.114708106	0.0557091	0.2431507	0.114708106	0.055709077	0.243150685
beverages, fermented and products	0	0	0	0	0	0	0	0	0
milk - (excluding excluding butter butter) and products	1.34939846	0.542761	3.030303	1.349398461	0.542761	3.030303	1.349398461	0.54276098	3.03030303
bananas and products	0	0	0	0	0	0	0	0	0
sugar cane and products	0	0	0	0	0	0	0	0	0
potatoes and products	0.51368421	0.230769	1.079392	0.513684211	0.2307692	1.0793919	0.513684211	0.230769231	1.079391892
poultry meat and products	4.84667574	3.377142	7.333896	4.846675737	3.377142	7.3338957	4.846675737	3.377142004	7.333895702
mutton & and goat meat and products	7.63102725	5.469575	11.84604	7.631027254	5.4695748	11.84604	7.631027254	5.469574838	11.8460403
oranges, mandarines and products	0	0	0.004541	0	0	0.0045411	0	0	0.004541095
soyabeans and products	2.51951228	1.008263	5.186834	2.519512281	1.0082627	5.1868337	2.519512281	1.008262704	5.186833656
bovine meat and products	4.94991676	3.647418	7.342657	4.949916759	3.6474185	7.3426573	4.949916759	3.647418478	7.342657343
sweeteners, other and products	0	0	0	0	0	0	0	0	0
pigmeat and products	6.95104895	4.787524	12.1205	6.951048951	4.7875244	12.120502	6.951048951	4.787524446	12.12050211
oilcrops oil, other and products	0	0	0	0	0	0	0	0	0
cocoa beans and products	0.0142654	0.008978	0.030435	0.014265403	0.0089777	0.0304348	0.014265403	0.008977695	0.030434783
sorghum and products	21.4589371	7.803594	42.34479	21.45893706	7.8035939	42.344785	21.45893706	7.803593947	42.34478529
plantains and products	0.41248689	0.252494	0.775873	0.412486888	0.2524936	0.7758731	0.412486888	0.252493614	0.775873077
meat, other and products	0.62295082	0.239742	1.623424	0.62295082	0.2397423	1.6234239	0.62295082	0.239742329	1.62342393
millet and products	0.4869938	0.216236	3.650205	0.486993797	0.216236	3.6502048	0.486993797	0.216236003	3.650204775
apples and products	0.08781818	0.0414	0.181084	0.087818182	0.0414	0.1810839	0.087818182	0.0414	0.181083869
beverages, alcoholic and products	0	0	0	0	0	0	0	0	0
coffee and products	0.13429319	0.05068	0.355932	0.134293194	0.0506796	0.3559322	0.134293194	0.050679612	0.355932203
wine and products	0.35343035	0.173346	0.638314	0.353430353	0.1733464	0.6383142	0.353430353	0.173346402	0.638314176
beer and products	0	0	0	0	0	0	0	0	0
pineapples and products	0.29688418	0.161147	0.488389	0.296884178	0.161147	0.4883889	0.296884178	0.161146952	0.488388924
roots, other and products	1.53473684	0.81	2.488827	1.534736842	0.81	2.4888268	1.534736842	0.81	2.488826816
butter, ghee and products	0.12909134	0.083107	0.201946	0.129091336	0.083107	0.2019464	0.129091336	0.083107037	0.201946386
honey and products	0.09368116	0.051759	0.170924	0.093681159	0.0517591	0.1709245	0.093681159	0.051759127	0.170924476
nuts and products	1.03580447	0.555964	2.117887	1.035804473	0.5559637	2.1178869	1.035804473	0.555963683	2.11788691

Supplementary Table 4. The percentage of population at risk of inadequacy based on the three apparent intake of Se scenarios: Baseline NCT, National (maize Se) NCT, and small area (maize Se) NCT.

	Baseline apparent Se intake		*National apparent Se intake*		*Small area apparent Se intake*	
	Percentage (95% Confidence Intervals (CI))					
Urban	28	(25-31)	68	(65-72)	67	(64-71)
Rural	42	(41-44)	87	(86-88)	82	(81-84)
Northern	44	(41-47)	81	(79-83)	82	(81-84)
Central	41	(39-43)	88	(86-89)	85	(83-86)
Southern	37	(35-39)	81	(79-82)	74	(72-76)
Balaka	38	(31-46)	85	(77-93)	81	(73-89)
Blantyre	29	(22-35)	78	(74-83)	73	(66-80)
Blantyre City	26	(20-32)	65	(58-71)	67	(60-74)
Chikwawa	32	(26-38)	65	(56-74)	30	(23-38)
Chiradzulu	38	(34-42)	83	(79-88)	84	(80-89)
Chitipa	46	(37-54)	90	(87-94)	93	(89-96)
Dedza	46	(40-52)	93	(90-96)	94	(91-97)
Dowa	48	(42-54)	91	(86-95)	90	(84-95)
Karonga	45	(38-53)	78	(74-82)	77	(73-80)
Kasungu	44	(35-53)	91	(88-94)	83	(75-90)
Likoma	28	(17-38)	63	(52-75)	63	(52-75)
Lilongwe	39	(34-45)	91	(88-94)	91	(88-94)
Lilongwe City	25	(21-30)	68	(63-74)	65	(60-70)
Machinga	44	(39-49)	89	(85-92)	86	(81-91)
Mangochi	40	(33-46)	91	(87-94)	90	(84-96)
Mchinji	40	(35-46)	87	(82-92)	83	(77-88)
Mulanje	50	(45-55)	86	(82-91)	79	(72-86)
Mwanza	44	(38-50)	87	(82-93)	76	(70-83)
Mzimba	41	(35-47)	84	(79-89)	84	(79-89)
Mzuzu City	35	(28-42)	68	(62-75)	71	(66-77)
Neno	37	(32-43)	85	(80-90)	73	(65-81)
Nkhatabay	57	(49-66)	84	(77-90)	86	(80-91)
Nkhotakota	62	(56-68)	91	(88-95)	92	(88-95)
Nsanje	35	(27-44)	74	(66-82)	23	(12-35)
Ntcheu	39	(33-46)	90	(85-96)	91	(87-96)
Ntchisi	51	(44-57)	93	(91-96)	93	(90-95)
Phalombe	48	(41-56)	89	(85-93)	84	(78-90)
Rumphi	41	(35-47)	86	(81-91)	88	(84-92)
Salima	43	(37-50)	93	(90-96)	71	(57-85)
Thyolo	40	(33-48)	82	(78-86)	73	(66-80)
Zomba	30	(25-34)	81	(76-86)	83	(78-88)
Zomba City	19	(15-22)	62	(56-67)	60	(54-65)



Supplementary Figure 1. Comparison of the apparent median Se intakes (in μg/AFE/day) estimated using the baseline NCT (yellow), national maize NCT (dark blue) and small area maize NCT (light blue) per each district and region in Malawi.

7 Discussion and Conclusions

7.1 Main findings

The main goal of this study was to explore the use of georeferenced data on food/crop mineral micronutrient composition, to improve the estimation of dietary mineral micronutrient intakes and deficiency risks in contexts with highly localised food systems, such as Malawi. First, the current data and metadata available for mineral micronutrients was assessed in 19 Food Composition Tables and databases (FCTs) and Nutrient Conversion Tables (NCTs) for countries in sub-Saharan Africa. The scoping review (Chapter 3) revealed disparities in the quantity and quality of both data and metadata. The highest data scarcity was found for Se and I, which affected more than 70% of the values in Malawi FCT (van Graan et al., 2019). Despite data disparities, high proportion of missing values were found in key food categories, such as Zn, Se and I content in ‘pelagic fish’ and ‘freshwater fish’. Similarly, the Se content in ‘maize and products’ which presented more than 75% of missing values across all the FCTs/ NCTs reviewed for use in sub-Saharan Africa. Maize is the main staple crop in the region, and thus, the lack of Se concentration data in maize is a problem, particularly, in contexts where this crop is one of the main sources of this micronutrient like in Malawi, Lesotho and Zambia among other countries (Ekpa et al., 2019; Galani et al., 2020; Joy, Kumssa et al., 2015; Kang et al., 2023).

Furthermore, when data on Se concentration in foods were available, the reporting systems and data structures precluded the identification of the true origin of the data points. This lack of transparency made the data available unreliable for use in quantifying dietary micronutrient intakes, as the composition of food/crop samples is spatially variable. This is particularly true for Se, including in staple crops, for which spatial patterns in crop composition are the main factor influencing population variation in dietary Se intakes in countries with highly localised food systems (Gashu et al., 2021; Ligowe et al., 2020; Mutonhodza et al., 2022).

Hence, an open science framework for compiling and reporting food composition data with spatial information was developed to help improved the efficiency and transparency of the development of NCTs for nutrition (Chapter 4). This framework was prepared following current international guidelines developed by the FAO/INFOODS and in collaboration with the FAO Food and Nutrition Division (ESN). The work was validated against a subset of the Global NCT for the Supply and Utilization Account collated by the ESN team. With this framework, the impact of quantity, quality and geographic location of the underlying composition data could be evaluated when developing FCTs/NCTs for assessing micronutrients intakes and risk of inadequacies. However, despite the framework providing an advancement in FCT/NCT reporting systems, the need for high quality, locally relevant maize Se concentration was unsolved, which was one of the problems identified by the scoping review.

Maize Se concentration at different sub-national aggregations (e.g. 10 km-60 km buffers, EA groups and district level) were produced for use in dietary assessment in Malawi. To the best of our knowledge, this is the first time that these types of spatial aggregation (i.e. based on the household location) have been produced, empirically compared and combined with household-level consumption data to estimate apparent Se intakes. The approach contrasts with methods that are typical in population micronutrient intakes estimation, where composition data are 'borrowed' from FCTs developed for other countries or regions, or where single values of Se composition are applied across national geographies, including for staple crops and highly consumed foods. These approaches are likely to produce unreliable estimates of dietary Se intake and may mask important subnational differences. Thus, recently collected georeferenced maize Se sampled data across Malawi was identified and compared with the spatial patterns in plasma Se concentration to identify the subnational aggregation level (i.e. 10 km-60 km buffers, EA groups and district level) of high resolution (~250 m^2) maize Se concentration data that would best explain plasma Se concentration in women (15-49 years old). The statistical models (in Chapter 5) revealed that maize Se concentration was associated with plasma Se concentration at all the sub-national aggregation levels, with a decreasing trend in the DIC and 95% Credible Intervals (CI) at smaller aggregation scales, suggesting a better explanatory performance. Interestingly, the only socio-demographic characteristic that showed a consistent, but small, association with plasma Se concentration was age.

Finally, the importance of small area aggregation of maize Se concentration data was demonstrated by the differences in the estimates of dietary Se intakes and inadequacies in different districts in Malawi. Three NCTs, which are summarised in Table 1 in Chapter 6, provided Se content in foods reported as consumed in the Malawi Fourth Integrated Household Survey, 2016-17 (Malawi 2016-17 IHS4) and were used to estimate the dietary Se intakes and the prevalence of population at risk of apparent inadequate intake in Malawi. The use of these NCTs, which exemplified differences in geographic origin and spatial resolution of Se data, revealed that the use of small area maize Se concentration data improve the dietary Se intakes estimates by providing refined and small-scale estimates which allowed for the identification of districts and hotspots at potentially high and low inadequacy risks in Malawi. The estimates of dietary Se inadequacy risk produced by the small area maize Se data were in line with the prevalence of Se deficiency in women (15-49 years old) based on plasma Se concentration, including spatial patterns of subnational differences (Phiri et al., 2019). A lack of subnational maize Se composition data, or failure to use this information appropriately, would potentially lead to large errors in assessing Se inadequacy risks at population level and potentially mis-informed nutrition policies. For example, the estimated prevalence of inadequate Se intakes in the baseline scenario, when using maize with high Se concentration mostly from Kenya FCT (2018), was 40% (95% CI 38-41%); in comparison, the estimated prevalence of inadequate intakes was 85% (95% CI 83-85%) when using Malawi maize composition data applied at the national level, and 79% (95% CI 78-80%) when using small area maize (Figure 3 in Chapter 6).

Based on the results of this thesis, in settings where food systems are highly localised, for instance, in context where staple crops are cultivated and consumed through short value chains (e.g. based on subsistence farming and/or local food markets), the spatial variation of mineral micronutrients should be evaluated. When the mineral micronutrient in staple crops show clustered spatial distributions, the use of national, single values may not be recommended when evaluating dietary intakes, particularly, at subnational level. This is because single national values would fail to account for the local mineral composition in foods, and hence, its impact in the dietary intakes of those populations. Hence, it is encouraged, when possible, to collect georeferenced samples of the main staple crops in countries which are highly reliant on these staples for their mineral micronutrient intakes.

7.2 Implications for public health nutrition

7.2.1 The need for better quality and higher resolution nutrition data

Over a decade of reports and journal series have acknowledged the need for better data to inform nutrition policies (Black et al., 2013; FAO & Intake, 2022; International Food Policy Research Institute, 2014; Piwoz et al., 2019; Popkin et al., 2020). These calls for action and agenda-setting reports have led to an increase in nutrition data collection and modelling activities, particularly aiming at closing certain nutrition-related data gaps, such as biomarkers, anthropometry and dietary assessment, and to inform evidence-based policies and public health nutrition programs (Brown et al., 2021; Coates et al., 2017; De Quadros et al., 2022; Micha et al., 2018; Miller et al., 2021; Osgood-Zimmerman et al., 2018). However, one aspect that it is often given less attention is the sparsity and low quality of the micronutrient data in FCTs. This is despite the importance of accurate food composition data for the reliable evaluation of dietary nutrient intakes and associated inadequacy risks (Bruyn et al., 2016; Grande et al., 2024; Karageorgou et al., 2024; Segovia de la Revilla et al., 2023).

Fortunately, attention to the role and importance of composition data is growing, leading to FCT updates and new developments. For instance, from the FCTs reviewed and compiled in Chapters 4 and 5, Australia (Food Standards Australia New Zealand, 2022), Denmark (DTU, 2024) Japan (MEXT, 2023), New Zealand (New Zealand Institute for Plant and Food Research Limited & Ministry of Health, 2024) and Norway (Norwegian Food Safety Authority, 2024) have recently updated their FCTs, whereas Ethiopia and the global uFish are currently in development (personal communication with authors). In addition, a focus on open data has been increasing within the food composition compilers and agencies, which would facilitate the maintenance, enhancement and the reproducibility of both FCTs and NCTs (New Zealand Institute for Plant and Food Research Limited & Ministry of Health, 2024, Norwegian Food Safety Authority, 2024). However, most of these developments are only carried out in FCTs for high-income countries, which are then used as secondary data in FCT compilation and dietary intake estimations in low- and middle-income countries. Thus, the open science framework for compiling reproducible FCTs and NCTs could be used to leverage the use of open data and open science principles in the update of new composition data by further automatising and enhancing the data and metadata quality while reducing compilation costs.

The importance of local primary food composition data is highlighted throughout this study, however, for certain micronutrients and foods, primary national level (one single point) data may not be sufficient as demonstrated in Chapter 5 and 6. This is because often national averages tend to mask disparities within countries, as exemplified in Chapter 6, where geographic location of the households was a key determinant of the Se apparent intakes and inadequacies, which could only be well characterised by the use of small area food composition data. Particularly, when staple crops, in this study maize, are the main source of mineral micronutrient in contexts with monotonous diets, which in turn, are frequently linked to the most vulnerable populations (Gatica-Domínguez et al., 2021; Kang et al., 2023). For instance, in Malawi, the energy contribution from cereals, mostly maize, was higher amongst the lowest wealth (Joy, Kumssa, et al., 2015). Similarly, rural women had presented lower plasma Se concentration, and tend to have low dietary diversity, particularly those in the lower wealth quintiles. One of the drivers for the less nutritious diets is limited access to markets (both physical and economical), which in turn, make them more reliant on locally grown foods, of which the mineral content will be highly dependent on the soil and environmental characteristics of their area.

Although vulnerabilities and inequalities related to food and nutrition are context specific, other studies have also highlighted inequalities related to the residence (rural and urban) and geographic location (i.e. certain areas of a country/neighbourhood may be neglected or geographically challenging) and the importance of spatially disaggregated nutrition data (HLPE, 2023; Kang et al., 2023). Hence, there is a need for granular data to evaluate, propose and monitor policies and actions that could improve dietary intakes of populations, particularly for the most vulnerable.

Finally, despite the overall need of higher spatial resolution of food composition data to better target interventions, the cost, in term of resources, expertise and money needed to produce these datasets in already budget constrained settings is prohibiting (Karageorgou et al., 2024). Therefore, the target should be the collection of accurate, analytical, subnational food composition data, for key food sources, at the minimal resolution needed in each context. This, in turn, would likely improve the feasibility and impact of the data collected and the derived estimates. Finally, awareness and advocacy of the importance and relevance of accurate food composition data should continue, particularly aiming at funders and governments of low- and middle-income countries (FAO & Intake, 2022).

7.2.2 Modelling approaches to improve nutrition data

The increase of computational power has led to the development of new modelling techniques for use in population health, including in public health nutrition. Several initiatives have been modelling the impact of micronutrient deficiencies in mortality and morbidities (Arndt et al., 2024; Black et al., 2008; Hess et al., 2021; Swinburn et al., 2019). Furthermore, nutrition-related datasets (e.g. micronutrients, dietary intakes, etc.) have been used in modelling approaches to define food-based recommendation and dietary guidelines (Ferguson et al., 2008), identify nutritionally vulnerable groups (Beal et al., 2024), evaluate cost-effectiveness interventions (Adams et al., 2022), and large-scale food fortification programs and initiatives (Tang et al., 2021). Similarly, some modelling approaches have been proposed for food composition data improvements. The use of machine-learning techniques was used to predict the nutrient content of branded foods in the US (Ma et al., 2021). Hicks colleagues (2019) used Bayesian modelling to estimate the nutrient content and contributions to global intakes from maritime fisheries. Similarly, we used high spatial resolution maize Se composition data to study different levels of aggregation for use in estimating dietary Se intakes (Chapter 5).

In the context of food composition data, and particularly in data-constrained settings, these models present opportunities to potentially improve the estimates of dietary nutrient intakes produced. Nevertheless, it is important to understand the limitations of modelling and Artificial Intelligence (AI)-derived outputs. For example, the amount of data needed, and the quality of the results are conditional on the quality of the underlying data, which often are based on imperfect data. Other frequently raised concerns are the validity, generalisability, transparency and reproducibility of the data, metadata and methods (Beal et al., 2021; Engle-Stone, 2024). This is particularly true for food composition data, where quality, availability and validity of the data has been raised numerous times as a limitation, and which is often used as underlying data of these models. Therefore, the use of our framework would increase the transparency of modelling efforts and highlight important uncertainties and data gaps. The models and data collection approaches used and proposed in Chapters 5 and 6 are fully transparent and reproducible, and could be expanded and combined with other modelling initiatives, to increase the accuracy of their results. By applying the open science frameworks in the collection, collation and preparation of the datasets (e.g. FCTs and NCTs), an increase of the harmonisation of the data for the dietary intakes and inadequacy estimations could be achieved (Karageorgou et al., 2024). This would allow:

a) increased accuracy of the results as more high-quality analytical data would be generated,

b) increased trustworthiness of the output generated with relation to nutrient estimates by allowing data users and practitioners to evaluate and model the data quality,

c) increased efficiency, as data gaps would be clearly identified and effort could be directed where more data, which has the potential to influence the final results, are needed,

d) increased interoperability with multiple datasets which could be used to respond to critical research and policy questions.

7.2.3 Biomarkers of Se status and dietary Se intakes

Plasma Se concentration is a good marker of Se status in population, and often, is perceived as more robust source of information than dietary data. Nevertheless, the use of dietary Se intakes can provide meaningful information regarding the aetiology of the Se deficiency, particularly when high spatial resolution composition data is paired with consumption data. For instance, in Chapter 5, variation in the plasma Se concentration in women (15-49 years old) in the Malawi MND-DHS (2015-2016) was not explained by individual socio-demographic and biological characteristics such as, inflammation status or wealth, whereas it was well characterised by sub-regional maize Se concentration, and thus, food intakes being the main driver of Se status (Fairweather-Tait et al., 2011; Hurst et al., 2013; Thomson, 2004).

The value of data triangulation has been recognised and proven by the increasing number of country-representative surveys that are collecting simultaneously micronutrient biomarker and dietary data (FAO & Intake, 2022; FGoN & IITA, 2024; Woldeyohannes et al., 2023). Although, biomarker and dietary data should ideally be collected repeatedly for the same sample of people, having high spatial resolution data for spatially depended micronutrients can contribute when drawing inferences from population level data (Belhadj et al., 2020; Combs, 2015; Fairweather-Tait et al., 2011; Thomson, 2004). For instance, in Chapter 6, the use of small area maize Se concentration identified the districts of Nsanje and Chikwawa as those with the least risk of Se inadequacy which was in line with the plasma Se status. Interestingly, Salima showed the second highest median plasma Se concentration in women (15-49 years old) (130 ng mL^{-1}, IQR 77.5-165) well above the deficiency threshold, which was not fully consistent with the apparent Se intakes: the median (IQR) apparent Se intake was 26.8 µg AFE^{-1} day^{-1}

(17.14-50.11) which was higher than the country median, but still well below the Estimated Average Requirement (EAR; for adult women, i.e. 45 µg day^{-1}). Some reasons for these discrepancies can be associated with limitations of the household data, as presented in the discussion section of Chapter 6. Others may be linked to variations in the micronutrients content in crops, i.e. Se driven by diverse soil types and environmental conditions or due to the lack of local representative Se data for fish, as reported in Chapter 3 (Segovia de la Revilla et al., 2023) and Chapter 5. This could have led to an overall underestimation of the dietary intakes of Se.

Furthermore, these findings are not conclusive as the representativeness of the Malawi MNS-DHS (2015-2016) survey only allows for regional disaggregation, and hence, some of these 'extreme' plasma Se concentration could be due to sampling error. For instance, the households sampled in the MNS-DHS could be located in a particularly high Se area within the Salima district. Nevertheless, in all the cases, the apparent intakes were better characterised by the use of maize NCT than by the national maize NCT. Hence, high spatial resolution Se concentration in foods paired with georeferenced food consumption information is key to identify potentially vulnerable groups, and to plan interventions at subnational levels, such as district level.

7.2.4 Policy implications of the research

In Malawi, the plasma Se and maize Se concentration showed consistent spatial association at all the subregion levels (Chapter 5). Hence, small area composition data may provide essential information for evaluating the population status, and for planning targeted, location specific interventions. The scenarios presented in Chapter 6 can be used to identify areas with low Se concentration in maize grain which are concurrent with high levels of Se deficiency. Then, using the methods presented in the same Chapter, the population that could potentially be reached by an increase in the Se concentration of maize grain could be modelled using different policy intervention scenarios, for example, agronomic biofortification (i.e. applying Se-enriched fertilisers to crops) or by increasing the maize production in regions with naturally high bioavailable Se.

Agronomic fortification has been a successful public health intervention to alleviate Se deficiencies in Finland. The government included Se as a mandatory composition of mineral fertilisers since 1984, which resulted in an overall increase in the Se concentration in crops, livestock and derived food products, and that translated into an increase of plasma Se concentration in the population (Alfthan et al., 2015). Relatedly, studies in the Eastern and Southern Africa regions have shown that maize Se concentration is highly responsive to Se application via fertilisers, and that consumption of Se-biofortified maize leads to large increases in plasma Se among women and children (Chilimba et al., 2014; Gondwe, 2018; Goredema-Matongera et al., 2021; Joy et al., 2021).

From the studies showing a positive relationship between Se uptake from fertilisers by crops and the potential effect on human Se status, one is particularly relevant: the pilot study of Joy and colleagues (2021) which demonstrated that agronomic biofortification of maize with Se could be a highly cost-effective intervention to address Se deficiency in Malawi. In the study, consumption of agronomically fortified *Granmil* flour led to a large increase in mean (standard deviation) plasma Se concentration among participants: from 57.6 (17.0) μg L^{-1} (n = 88) to 107.9 (16.4) μg L^{-1} (n = 88) among women (20-45 years old) and from 46.4 (14.8) μg L^{-1} (n = 86) to 97.1 (16.0) μg L^{-1} (n = 88) among School Aged Children. The Se concentration in *Granmil* flour was 14.8 μg/100 g Fresh Weight Edible Portion (FW EP) which was around 10-fold higher than the median national (single value) maize flour (refined) in Malawi (1.1 μg/100 g FW EP) and substantially lower than the median values in some of the EAs in Nsanje and Chikwawa districts presented in the Chapter 6 (31.3 and 19.7 μg/100 g FW EP, respectively). This evidence suggests that agronomic fortification could be a potential solution to curb Se deficiency in Malawi, where the majority of cropland areas have soils with low plant-available Se concentrations, and hence, applying Se-rich fertiliser could increase not only maize, but other crops Se concentration (Chilimba et al., 2014; Muleya et al., 2021). Furthermore, the maps presented here can also guide policy makers in the identification of areas where Se-rich fertilisers will not be required. For instance, as described above, in the southern districts of Nsanje and Chikwawa, where Se concentration in maize, and other staple grains are already high while Se deficiency and apparent Se intakes were analogously low. This would increase the cost-efficiency of the intervention. The same model could be applied to other countries in the region which have similarly low bioavailable Se soils and highly localised food systems (Ligowe et al., 2020).

Another use of the data provided here is to identify vulnerable groups, particularly, when staple crops are the main source of mineral micronutrient in contexts with monotonous and less diverse diets, which in turn, are frequently linked to the most vulnerable populations. For instance, households on the lower wealth quantiles were less likely to meet minimum dietary diversity (MDD) scores in low- and middle-income countries and in Southern and Eastern Africa (Gatica-Domínguez et al., 2021; Kang et al., 2023). Similarly, in Malawi, the energy contribution from cereals, mostly maize, was higher amongst the lowest wealth quantiles and particularly high (>70%) for those living in urban areas (Joy, Kumssa, et al., 2015). These population may be more reliant on these staple crops for their micronutrient intakes, and thus, more susceptible to changes in the micronutrient content of these crops. Interestingly, our models in Chapter 5 showed, when accounting for spatial variation in maize, wealth index was not associated with plasma Se concentration in women (15-49 years old). As such, in Malawi, the mineral intakes, for which maize is their main source, would highly depend on the mineral content on locally grown maize, which in turn will be highly dependent on the soil and environmental characteristics of the area, and there may be little dependence on household wealth. Hence, there is a need for more granular food composition data to evaluate, implement and monitor policies and actions that would improve dietary intakes of populations, particularly for the most vulnerable.

7.3 Strengths and limitations

7.3.1 Strengths

One strength of this research is the use of open science approaches and transparent methods including for the development of FCTs/NCTs which was designed with key actors in the field of food composition data (e.g. FAO/INFOODS). These approaches were applied across all the chapters of this thesis, including publishing all the R scripts used in each chapter analyses as GitHub repositories for reproducibility and transparency. In addition, to the best of our knowledge, this is the first time that a geospatial method like INLA has been applied to inform the level of aggregation of high-resolution composition data for nutrition assessment. Finally, the work demonstrates the use of nationally representative crop composition surveys for assessing micronutrient intakes and inadequacy risks at national and sub-national scales, and the importance of using high resolution food composition data and small area estimates to appropriately inform public health nutrition policies.

7.3.2 Limitations

The main limitation is related to the transferability and applicability of the results for other nutrients and other settings beyond Se and Malawi. Selenium was considered an appropriate case study nutrient due to the responsiveness of the biomarker (i.e. plasma Se) to dietary Se intakes, hence its value for identifying the suitable aggregation level of georeferenced crop composition data. However, the approaches presented here could be followed to define appropriate crop composition aggregations in other context with localised food systems. The approach would be particularly useful in contexts with limited dietary diversity, including in Malawi, Ethiopia, Kenya and Zimbabwe, where staple crops drive the spatial variability in dietary micronutrient intakes, like Ca, Zn and Fe (Galani et al., 2020; Gashu et al., 2021; Ngigi et al., 2019). This may not be the case in other countries where staple crops may vary across regions and/or seasons, or for micronutrients where spatial structure in crop composition is low or inexistent at scales relevant for food systems.

Finally, another limitation is the lack of large-scale crop composition datasets in other settings and the potentially large costs associated with generating such data. The GeoNutrition data used in the current studies are limited to Ethiopia and Malawi, and thus, the potential reusability of the methods developed here with smaller resolution datasets or number of georeferenced samples is yet to be explored. Despite that, results from our analyses showed that aggregating crop composition data at district level, which is likely to be less resource intensive, would still perform substantially better than regional or country-level estimates for reliable estimation of Se inadequacy risks.

7.4 Way forward

The results of this research highlighted the need for collecting higher quality and resolution food composition data, particularly for mineral micronutrients such as Se, in contexts with localised food systems. Crop/food sample collection, analysis and reporting could be a combined effort between the nutrition and food science community and other research fields such as agronomy, agriculture and soil science. For instance, a number of agriculture-led projects have been collecting data on the mineral content (and other nutrients) in crops when testing different agronomic practices which could be used to populate food composition databases. However, the disconnection between the fields has led to poor reusability of the data produced by agriculture research and agronomy in nutrition sciences. One of the most common issues that impede the reusability of agriculture data in FCTs is that water content of the samples is often not collected. In nutrition, the water content of the food is important as foods are reported as consumed (e.g. 'edible portion on fresh weight basis'). For instance, if a person reported consuming 100 g of lettuce, the water content of that food is needed to allocate the correct amount of nutrient consumed by the person. Although this may be a lesser issue for 'dried' foods, such as grains, flours, etc. as the water content is rather constant in those foods, nonetheless the lack of moisture content data may condition the reusability of the data in FCTs, as water content is used to calculate other components, such as carbohydrates by difference, to complete the nutrients reported (e.g. for imputing values of proteins or vitamins from other sources), and is used in quality checks (e.g. sum of proximate) (FAO/INFOODS, 2012).

In addition to water content of foods, when new datasets are collated aiming to be reused by nutrition and food scientists, other information requirements should be: time and location of the samples, and sampling method and the use of standardised analytical methods. For instance, in a recent study where minerals in staple crops were collected in Ethiopia, information on the location of the samples were not available despite the importance of the geographic location for Se described by the same authors (Godebo et al., 2023; Nyachoti et al., 2021).

Equally important would be the use of adequate data structures and standardised metadata which allow users and machines to access and understand the information. By doing so, new modelling approaches to food composition could be developed and tested, including the combination with other large-scale datasets (e.g. satellite image) to build predictive models. This could be achieved by building on the Findable, Accessible, Interoperable and Reusable (FAIR) principles and setting Minimum Reporting Standards (MRS) and could help both re-usability and findability which, in turn, would increase the cost-effectiveness and impact of the research. Additionally, it would help inform better nutrition interventions.

Despite the unique responsiveness of the plasma Se to dietary intakes, there is evidence of the spatial relationship between other mineral micronutrients, dietary intakes and biomarkers in Ethiopia, Malawi, and elsewhere (Belay et al., 2021, 2022; Botoman et al., 2022; Gondwe, 2018; Phiri et al., 2021). For example, in Ethiopia, the spatial pattern of Zn supplies from teff and wheat showed similarities with the spatial pattern of Zn deficiencies (Belay et al., 2021; Gashu et al., 2021). This evidence has led to the inclusion of micronutrient (Zn, Fe, Cu and Mg) soil and crop sampling in the latest Ethiopian National Food and Nutrition Survey (Woldeyohannes et al., 2023). Although further research on the spatial relationship between these biomarkers and staple crops should be developed, the use of high-resolution composition data will likely to produce more accurate estimates, particularly at small scales (<45 km). To this extent, and in lieu of a better approach, we recommend using small area (e.g. based on the household EAs) for other mineral micronutrients, particularly for crops that are a source of multiple micronutrients. For instance, using small aggregations (i.e. district level or based on EAs) estimates for Zn, Ca and Fe in maize in Malawi, Ethiopia, certain areas of Kenya, Tanzania or Zambia (Galani et al., 2020; Nguyen & Qaim, 2025).

Here we presented a case study for a mineral micronutrient with spatial variation, however, there may be other nutrients that may have different variation patterns. For example, Grande and colleagues (2016) found disparities in the results when evaluating vitamin A supplies from the Brazilian Household Budget Survey (HBS) (2008–2009) and three FCTs. Additionally, vitamin B12 has been shown to vary through time (including seasonal variation) in fish, which may be due to feeding patterns, weather conditions or other factors yet to be studied (Rego et al., 2022). These patterns are likely to influence the vitamin B12 intakes of population that may rely on a limited number of foods to reach their nutrient intakes.

Finally, we hope to see an increase in the mobilisation of resources and investment from funders and governments to collect high resolution food composition data and more spatially disaggregated (i.e. providing GPS coordinates at cluster level) dietary intake data. Some lessons that could be learnt from other dietary data collection initiatives, such as the need for diverse stakeholders to recognise the importance of food composition data for their goal and interest in order to drive agendas (FAO & Intake, 2022). For instance, the importance of accurate food composition data to identify nutrient gaps and implement and monitor the effectiveness of agricultural policies, such as fertiliser subsidies and/or fortification initiatives. Similarly, government officials in the trade department may identify and monitor changes in the dietary intakes due to the import of crop/commodities from certain countries. The health department and the education ministry may be interested in knowing the nutrient content of the food served as part of their school meal programme to assess the impact on educational outcomes. Hence, educating and advocating for the role of food composition data in all these policy decisions may be the only way to increase the much-needed attention and investment in the field of food composition.

7.5 Conclusion

In conclusion, high-quality food composition data is vital for the adequate assessment of micronutrient inadequacies, particularly in context with high risk of deficiencies and localised and monotonous diets. Furthermore, for micronutrients that present spatial variability, the use of sub-national food composition data are recommended to avoid masking potential high vs low intakes areas. Finally, improved reporting systems and structures for the nutrient content in foods are needed to: a) assess uncertainty and trustworthiness of the micronutrient inadequacy estimates, b) identify data gaps that may bias the estimates, and when possible, c) the use of modelling techniques to improve the estimates.

References

Adams, K. P., Luo, H., Vosti, S. A., Kagin, J., Ngnie-Teta, I., Ndjebayi, A., Assiene, J. G., & Engle-Stone, R. (2022). Comparing estimated cost-effectiveness of micronutrient intervention programs using primary and secondary data: Evidence from Cameroon. *Annals of the New York Academy of Sciences*, *1510*(1), 100–120. https://doi.org/10.1111/nyas.14726

Alfthan, G., Eurola, M., Ekholm, P., Venäläinen, E.-R., Root, T., Korkalainen, K., Hartikainen, H., Salminen, P., Hietaniemi, V., Aspila, P., & Aro, A. (2015). Effects of nationwide addition of selenium to fertilizers on foods, and animal and human health in Finland: From deficiency to optimal selenium status of the population. *Journal of Trace Elements in Medicine and Biology*, *31*, 142–147. https://doi.org/10.1016/j.jtemb.2014.04.009

Arndt, M. B., Abate, Y. H., Abbasi-Kangevari, M., ElHafeez, S. A., Abdelmasseh, M., Abd-Elsalam, S., Abdulah, D. M., Abdulkader, R. S., Abidi, H., Abiodun, O., Aboagye, R. G., Abolhassani, H., Abtew, Y. D., Abu-Gharbieh, E., Abu-Rmeileh, N. M., Acuna, J. M., Adamu, K., Adane, D. E., Addo, I. Y., … Reiner, R. C. (2024). Global, regional, and national progress towards the 2030 global nutrition targets and forecasts to 2050: A systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, *0*(0). https://doi.org/10.1016/S0140-6736(24)01821-X

Beal, T., Herforth, A., Sundberg, S., Hess, S. Y., & Neufeld, L. M. (2021). Differences in modelled estimates of global dietary intake. *The Lancet*, *397*(10286), 1708–1709. https://doi.org/10.1016/S0140-6736(21)00714-5

Beal, T., Manohar, S., Miachon, L., & Fanzo, J. (2024). Nutrient-dense foods and diverse diets are important for ensuring adequate nutrition across the life course. *Proceedings of the National Academy of Sciences*, *121*(50), e2319007121. https://doi.org/10.1073/pnas.2319007121

Belay, A., Gashu, D., Joy, E. J. M., Lark, M. R., Chagumaira, C., Zerfu, D., Ander, L. E., Young, S. D., Bailey, E. H., & Broadley, M. R. (2022). Mineral micronutrient status and spatial distribution among the Ethiopian population. *British Journal of Nutrition*, 1–11. https://doi.org/10.1017/S0007114522000319

Belay, A., Gashu, D., Joy, E. J. M., Lark, R. M., Chagumaira, C., Likoswe, B. H., Zerfu, D., Ander, E. L., Young, S. D., Bailey, E. H., & Broadley, M. R. (2021). Zinc deficiency is highly prevalent and spatially dependent over short distances in Ethiopia. *Scientific Reports*, *11*(1), 6510. https://doi.org/10.1038/s41598-021-85977-x

Belhadj, M., Kazi Tani, L. S., Dennouni Medjati, N., Harek, Y., Dali Sahi, M., Sun, Q., Heller, R., Behar, A., Charlet, L., & Schomburg, L. (2020). Se Status Prediction by Food Intake as Compared to Circulating Biomarkers in a West Algerian Population. *Nutrients*, *12*(12), Article 12. https://doi.org/10.3390/nu12123599

Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., de Onis, M., Ezzati, M., Mathers, C., & Rivera, J. (2008). Maternal and child undernutrition: Global and regional exposures and health consequences. *The Lancet*, *371*(9608), 243–260. https://doi.org/10.1016/S0140-6736(07)61690-0

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R., & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*, *382*(9890), 427–451. https://doi.org/10.1016/S0140-6736(13)60937-X

Botoman, L., Chagumaira, C., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Chimungu, J. G., Gameda, S., Gashu, D., Haefele, S. M., Joy, E. J. M., Kumssa, D. B., Ligowe, I. S., McGrath, S. P., Milne, A. E., Munthali, M., Towett, E., Walsh, M. G., Wilson, L., … Nalivata, P. C. (2022). Soil and landscape factors influence geospatial variation in maize grain zinc concentration in Malawi. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-12014-w

Brown, K. H., Moore, S. E., Hess, S. Y., McDonald, C. M., Jones, K. S., Meadows, S. R., Manger, M. S., Coates, J., Alayon, S., & Osendarp, S. J. M. (2021). Increasing the availability and utilization of reliable data on population micronutrient (MN) status globally: The MN Data Generation Initiative. *The American Journal of Clinical Nutrition*, *nqab173*. https://doi.org/10.1093/ajcn/nqab173

Bruyn, J. de, Ferguson, E., Allman-Farinelli, M., Darnton-Hill, I., Maulaga, W., Msuya, J., Alders, R., de Bruyn, J., Ferguson, E., Allman-Farinelli, M., Darnton-Hill, I., Maulaga, W., Msuya, J., & Alders, R. (2016). Food composition tables in resource-poor settings: Exploring current limitations and opportunities, with a focus on animal-source foods in sub-Saharan Africa. *The British Journal of Nutrition*, *116*(10), 1709–1719. https://doi.org/10.1017/S0007114516003706

Chilimba, A. D. C., Young, S. D., & Joy, E. J. M. (2014). *Agronomic biofortification of maize, soybean and groundnut with selenium in intercropping and sole cropping systems*. https://doi.org/10.5897/AJAR2014.8978

Coates, J. C., Colaiezzi, B. A., Bell, W., Charrondiere, U. R., & Leclercq, C. (2017). Overcoming Dietary Assessment Challenges in Low-Income Countries: Technological Solutions Proposed by the International Dietary Data Expansion (INDDEX) Project. *Nutrients*, *9*(3), 289. https://doi.org/10.3390/nu9030289

Combs, G. F. (2015). Biomarkers of Selenium Status. *Nutrients*, *7*(4), Article 4. https://doi.org/10.3390/nu7042209

De Quadros, V. P., Balcerzak, A., Allemand, P., De Sousa, R. F., Bevere, T., Arsenault, J., Deitchler, M., & Holmes, B. A. (2022). Global Trends in the Availability of Dietary Data in Low and Middle-Income Countries. *Nutrients*, *14*(14), 2987. https://doi.org/10.3390/nu14142987

DTU (Technical University of Denmark) Food Institute, 2024. Frida Food Databank. Version 5.2. Copenhagen, DTU, National Food Institute. [Cited 07 January 2025]. https://frida.fooddata.dk.

Ekpa, O., Palacios-Rojas, N., Kruseman, G., Fogliano, V., & Linnemann, A. R. (2019). Sub-Saharan African Maize-Based Foods—Processing Practices, Challenges and Opportunities. *Food Reviews International*, *35*(7), 609–639. https://doi.org/10.1080/87559129.2019.1588290

Engle-Stone, R. (2024). Artificial intelligence-assisted mobile dietary assessment: Time to expand the standard toolkit? *The American Journal of Clinical Nutrition*, *120*(5), 1005–1006. https://doi.org/10.1016/j.ajcnut.2024.09.008

Fairweather-Tait, S. J., Bao, Y., Broadley, M. R., Collings, R., Ford, D., Hesketh, J. E., & Hurst, R. (2011). Selenium in Human Health and Disease. *Antioxidants & Redox Signaling*, *14*(7), 1337–1383. https://doi.org/10.1089/ars.2010.3275

FAO & Intake. (2022). *Global report on the state of dietary data*. FAO. https://doi.org/10.4060/cb8679en

FAO/INFOODS. (2012). *FAO/INFOODS Guidelines for Checking Food Composition Data prior to the Publication of a User Table/Database-Version 1.0*. FAO. https://www.fao.org/3/ap810e/ap810e.pdf

FGoN (Federal Government of Nigeria) & IITA (International Institute of Tropical Agriculture). (2024). *National Food Consumption and Micronutrient Survey 2021. Key Findings.* FGoN and IITA. https://www.iita.org/wp-content/uploads/2024/05/NFCMS-2021-Key-Findings.pdf

Ferguson, E. L., Briend, A., Darmon, N., E.L., F., & A., B. (2008). Can optimal combinations of local foods achieve the nutrient density of the F100 catch-up diet for severe malnutrition? *Journal of Pediatric Gastroenterology and Nutrition*, *46*(4), 447–452. https://doi.org/10.1097/MPG.0b013e318156cf5c

Food Standards Australia New Zealand. (2022). *Australian Food Composition Database – Release 2*. FSANZ. https://www.foodstandards.gov.au/science-data/monitoringnutrients/afcd/australian-food-composition-database-download-excel-files

Galani, Y. J. H., Orfila, C., & Gong, Y. Y. (2020). A review of micronutrient deficiencies and analysis of maize contribution to nutrient requirements of women and children in Eastern and Southern Africa. *Critical Reviews in Food Science and Nutrition*, *0*(0), 1–24. https://doi.org/10.1080/10408398.2020.1844636

Gashu, D., Nalivata, P. C., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A. E., Mossa, A. W., Munthali, M., … Broadley, M. R. (2021). The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*, *594*(7861), Article 7861. https://doi.org/10.1038/s41586-021-03559-3

Gatica-Domínguez, G., Neves, P. A. R., Barros, A. J. D., & Victora, C. G. (2021). Complementary Feeding Practices in 80 Low- and Middle-Income Countries: Prevalence of and Socioeconomic Inequalities in Dietary Diversity, Meal Frequency, and Dietary Adequacy. *The Journal of Nutrition*, *151*(7), 1956–1964. https://doi.org/10.1093/jn/nxab088

Godebo, T. R., Stoner, H., Kodsup, P., Stoltzfus, M., Nyachoti, S., Atkins, S., & Jeuland, M. (2023). Selenium in drinking water and cereal grains, and biomarkers of Se status in urine and fingernails of the Main Ethiopian Rift Valley population. *Journal of Trace Elements in Medicine and Biology*, *77*, 127137. https://doi.org/10.1016/j.jtemb.2023.127137

Gondwe, C. (2018). *Improving Yields and Enhancing Dietary Zinc and Selenium Intake in the Zambian Population Through Agronomic Biofortification of Maize And Wheat*. https://www.semanticscholar.org/paper/Improving-Yields-and-Enhancing-Dietary-Zinc-and-in-Gondwe/ececa5f77316642e81601ebeaa99e6f4fabebd0c

Goredema-Matongera, N., Ndhlela, T., Magorokosho, C., Kamutando, C. N., van Biljon, A., & Labuschagne, M. (2021). Multinutrient Biofortification of Maize (Zea mays L.) in Africa: Current Status, Opportunities and Limitations. *Nutrients*, *13*(3), Article 3. https://doi.org/10.3390/nu13031039

Grande, F., Giuntini, E. B., Lajolo, F. M., & de Menezes, E. W. (2016). How do calculation method and food data source affect estimates of vitamin A content in foods and dietary intake? *Journal of Food Composition and Analysis*, *46*, 60–69. https://doi.org/10.1016/j.jfca.2015.11.006

Grande, F., Ueda, Y., Masangwi, S., & Holmes, B. (2024). *Global nutrient conversion table for FAO supply utilization accounts*. FAO. https://doi.org/10.4060/cc9678en

Hess, S. Y., McLain, A. C., Lescinsky, H., Brown, K. H., Afshin, A., Atkin, R., & Osendarp, S. J. (2021). Basis for changes in the disease burden estimates related to vitamin A and zinc deficiencies in the 2017 and 2019 Global Burden of Disease Studies. *Public Health Nutrition*, 1–7. https://doi.org/10.1017/S1368980021004821

Hicks, C. C., Cohen, P. J., Graham, N. A. J., Nash, K. L., Allison, E. H., D’Lima, C., Mills, D. J., Roscher, M., Thilsted, S. H., Thorne-Lyman, A. L., & MacNeil, M. A. (2019). Harnessing global fisheries to tackle micronutrient deficiencies. *Nature*, *574*(7776), Article 7776. https://doi.org/10.1038/s41586-019-1592-6

HLPE. (2023). *Reducing inequalities for food security and nutrition*. CFS HLPE-FSN.

Hurst, R., Siyame, E. W. P., Young, S. D., Chilimba, A. D. C., Joy, E. J. M., Black, C. R., Ander, E. L., Watts, M. J., Chilima, B., Gondwe, J., Kang'ombe, D., Stein, A. J., Fairweather-Tait, S. J., Gibson, R. S., Kalimbira, A. A., & Broadley, M. R. (2013). Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi. *Scientific Reports*, *3*, 1425–1425. Africa-Wide Information. https://doi.org/10.1038/srep01425

International Food Policy Research Institute. (2014). *Global Nutrition Report* (0 ed.). International Food Policy Research Institute. https://doi.org/10.2499/9780896295643

Joy, E. J. M., Kalimbira, A. A., Sturgess, J., Banda, L., Chiutsi-Phiri, G., Manase, H., Gondwe, J., Ferguson, E. L., Kalumikiza, Z., Bailey, E. H., Young, S. D., Matandika, L., Mfutso-Bengo, J., Millar, K., Niksic, M., Segovia de la Revilla, L., Likoswe, B. H., Phuka, J. C., Phiri, F. P., … Allen, E. (2021). Biofortified Maize Improves Selenium Status of Women and Children in a Rural Community in Malawi: Results of the Addressing Hidden Hunger With Agronomy Randomized Controlled Trial. *Frontiers in Nutrition*, *8*, 788096. https://doi.org/10.3389/fnut.2021.788096

Joy, E. J. M., Kumssa, D. B., Broadley, M. R., Watts, M. J., Young, S. D., Chilimba, A. D. C., & Ander, E. L. (2015). Dietary mineral supplies in Malawi: Spatial and socioeconomic assessment. *BMC Nutrition*, *1*(1), 42. https://doi.org/10.1186/s40795-015-0036-4

Kang, Y., Heidkamp, R. A., Mako-Mushaninga, K., Garg, A., Matji, J. N., Nyawo, M., Craig, H. C., & Thorne-Lyman, A. L. (2023). Factors associated with diet diversity among infants and young children in the Eastern and Southern Africa region. *Maternal & Child Nutrition*, *19*(3), e13487. https://doi.org/10.1111/mcn.13487

Karageorgou, D., Lara-Castor, L., Padula De Quadros, V., Ferreira De Sousa, R., Holmes, B. A., Ioannidou, S., Mozaffarian, D., & Micha, R. (2024). Harmonizing dietary datasets for global surveillance: Methods and findings from the Global Dietary Database. *Public Health Nutrition*, 1–52. https://doi.org/10.1017/S1368980024000211

Ligowe, I. S., Phiri, F. P., Ander, E. L., Bailey, E. H., Chilimba, A. D. C., Gashu, D., Joy, E. J. M., Lark, R. M., Kabambe, V., Kalimbira, A. A., Kumssa, D. B., Nalivata, P. C., Young, S. D., & Broadley, M. R. (2020). Selenium deficiency risks in sub-Saharan African food systems and their geospatial linkages. *Proceedings of the Nutrition Society*, *79*(4), 457–467. https://doi.org/10.1017/S0029665120006904

Ma, P., Li, A., Yu, N., Li, Y., Bahadur, R., Wang, Q., & Ahuja, J. K. (2021). Application of machine learning for estimating label nutrients using USDA Global Branded Food Products Database, (BFPD). *Journal of Food Composition and Analysis*, *100*, 103857. https://doi.org/10.1016/j.jfca.2021.103857

MEXT (Ministry of Education, Culture, Sports, Science and Technology). 2023 Supplement to the Standard Tables of Food Composition in Japan (8th Edition). https://www.mext.go.jp/a_menu/syokuhinseibun/mext_00001.html (only in Japanese)

Micha, R., Coates, J., Leclercq, C., Charrondiere, U. R., & Mozaffarian, D. (2018). Global Dietary Surveillance: Data Gaps and Challenges. *Food and Nutrition Bulletin*, *39*(2), 175–205. https://doi.org/10.1177/0379572117752986

Miller, V., Singh, G. M., Onopa, J., Reedy, J., Shi, P., Zhang, J., Tahira, A., Morris, M. L. S., Marsden, D. P., Kranz, S., Stoyell, S., Webb, P., Micha, R., & Mozaffarian, D. (2021). Global Dietary Database 2017: Data availability and gaps on 54 major foods, beverages and nutrients among 5.6 million children and adults from 1220 surveys worldwide. *BMJ Global Health*, *6*(2), e003585. https://doi.org/10.1136/bmjgh-2020-003585

Muleya, M., Young, S. D., Reina, S. V., Ligowe, I. S., Broadley, M. R., Joy, E. J. M., Chopera, P., & Bailey, E. H. (2021). Selenium speciation and bioaccessibility in Se-fertilised crops of dietary importance in Malawi. *Journal of Food Composition and Analysis*, *98*, 103841. https://doi.org/10.1016/j.jfca.2021.103841

Mutonhodza, B., Joy, E. J. M., Bailey, E. H., Lark, M. R., Kangara, M. G. M., Broadley, M. R., Matsungo, T. M., & Chopera, P. (2022). Linkages between soil, crop, livestock, and human selenium status in Sub-Saharan Africa: A scoping review. *International Journal of Food Science & Technology*, *57*(10), 6336–6349. https://doi.org/10.1111/ijfs.15979

New Zealand Institute for Plant and Food Research Limited & Ministry of Health. 2024. New Zealand FOODfiles™ 2024 (Version 01). The New Zealand Food Composition Database. Auckland. https://www.foodcomposition.co.nz/foodfiles/

Ngigi, P., Lachat, C., Masinde, P., & Laing, G. (2019). Agronomic biofortification of maize and beans in Kenya through selenium fertilization. *Environmental Geochemistry and Health*, *41*. https://doi.org/10.1007/s10653-019-00309-3

Nguyen, T.-T., & Qaim, M. (2025). Local and regional food production diversity are positively associated with household dietary diversity in rural Africa. *Nature Food*, *6*(2), 205–212. https://doi.org/10.1038/s43016-024-01096-6

Norwegian Food Safety Authority, 2024. Norwegian Food Composition Database. Oslo. https://www.matvaretabellen.no.

Nyachoti, S., Adebayo, S., & Godebo, T. R. (2021). Elemental composition of teff (a gluten-free grain), maize and wheat: Staple crops in the Main Ethiopian Rift Valley. *Journal of Food Composition and Analysis*, *100*, 103660. https://doi.org/10.1016/j.jfca.2020.103660

Osgood-Zimmerman, A., Millear, A. I., Stubbs, R. W., Shields, C., Pickering, B. V., Earl, L., Graetz, N., Kinyoki, D. K., Ray, S. E., Bhatt, S., Browne, A. J., Burstein, R., Cameron, E., Casey, D. C., Deshpande, A., Fullman, N., Gething, P. W., Gibson, H. S., Henry, N. J., … Hay, S. I. (2018). Mapping child growth failure in Africa between 2000 and 2015. *Nature*, *555*(7694), 41–47. https://doi.org/10.1038/nature25760

Phiri, F. P., Ander, E. L., Bailey, E. H., Chilima, B., Chilimba, A. D. C., Gondwe, J., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Phuka, J. C., Salter, A., Suchdev, P. S., Watts, M. J., Young, S. D., & Broadley, M. R. (2019a). The risk of selenium deficiency in Malawi is large and varies over multiple spatial scales. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-43013-z

Phiri, F. P., Ander, E. L., Lark, R. M., Joy, E. J. M., Kalimbira, A. A., Suchdev, P. S., Gondwe, J., Hamilton, E. M., Watts, M. J., & Broadley, M. R. (2021). Spatial analysis of urine zinc (Zn) concentration for women of reproductive age and school age children in Malawi. *Environmental Geochemistry and Health*, *43*(1), 259–271. https://doi.org/10.1007/s10653-020-00700-5

Piwoz, E., Rawat, r., Fracassi, P., & David K. (2019). Strengthening the Nutrition Data Value Chain for Accountability and Action. *Data in Nutrition- Magazine -2019*, *2019*(1). https://doi.org/10.52439/HDXT8911

Popkin, B. M., Corvalan, C., & Grummer-Strawn, L. M. (2020). Dynamics of the double burden of malnutrition and the changing nutrition reality. *The Lancet*, *395*(10217), 65–74. https://doi.org/10.1016/S0140-6736(19)32497-3

Rego, A., Coelho, I., Motta, C., Cardoso, C., Gomes-Bispo, A., Afonso, C., Prates, J. A. M., Bandarra, N. M., Silva, J. A. L., & Castanheira, I. (2022). Seasonal variation of chub mackerel (Scomber colias) selenium and vitamin B12 content and its potential role in human health. *Journal of Food Composition and Analysis*, *109*, 104502. https://doi.org/10.1016/j.jfca.2022.104502

Segovia de la Revilla, L., Ferguson, E. L., Dooley, C., Osman, G., Ander, E. L., & Joy, E. J. M. (2023). The availability and geographic location of open-source food composition data used to estimate micronutrient intakes in sub-Saharan Africa: A scoping review. *Journal of Food Composition and Analysis*, *120*, 105322. https://doi.org/10.1016/j.jfca.2023.105322

Swinburn, B. A., Kraak, V. I., Allender, S., Atkins, V. J., Baker, P. I., Bogard, J. R., Brinsden, H., Calvillo, A., Schutter, O. D., Devarajan, R., Ezzati, M., Friel, S., Goenka, S., Hammond, R. A., Hastings, G., Hawkes, C., Herrero, M., Hovmand, P. S., Howden, M., … Dietz, W. H. (2019). The Global Syndemic of Obesity, Undernutrition, and Climate Change: The Lancet Commission report. *The Lancet*, *393*(10173), 791–846. https://doi.org/10.1016/S0140-6736(18)32822-8

Tang, K., Adams, K. P., Ferguson, E. L., Woldt, M., Kalimbira, A. A., Likoswe, B., Yourkavitch, J., Chrisinger, B., Pedersen, S., Segovia De La Revilla, L., Dary, O., Ander, E. L., & Joy, E. J. M. (2021). Modeling food fortification contributions to micronutrient requirements in Malawi using Household Consumption and Expenditure Surveys. *Annals of the New York Academy of Sciences*, *n/a*(n/a). https://doi.org/10.1111/nyas.14697

Thomson, C. D. (2004). Assessment of requirements for selenium and adequacy of selenium status: A review. *European Journal of Clinical Nutrition*, *58*(3), 391–402. https://doi.org/10.1038/sj.ejcn.1601800

van Graan, A., Chetty, J., Jumat, M., Masangwi, S., Mwangwela, A., Phiri, F. P., Ausman, L. M., Ghosh, S., & Marino-Costello, E. (Eds.). (2019). *Malawian Food Composition Tables* (1st ed.).

Woldeyohannes, M., Girma, M., Petros, A., Hussen, A., Samuel, A., Dinssa, D. A., Challa, F., Laillou, A., Chitekwe, S., Baye, K., Noor, R., Donze, A. S., Tollera, G., Dangiso, M. H., Tadesse, L., Zelalem, M., & Tessema, M. (2023). Ethiopia National Food and Nutrition Survey to inform the Ethiopian National Food and Nutrition Strategy: A study protocol. *BMJ Open*, *13*(4), e067641. https://doi.org/10.1136/bmjopen-2022-067641

List of Annexes

Annex 1: Data Management Plan for Research Students (v.2.4.0).

Annex 2: Ethical approval letter.

Annex 3: Certificate of completion: Research Ethics training.

Annex 4: List of publications and presentations

Data Management Plan for Research Students

Project title	Geospatial analysis of food composition data to estimate population micronutrient intake
Author name	Lucia Segovia de la Revilla
Supervisor	Dr Edward Joy, Dr Elaine Ferguson, Dr Claire Dooley
Contact email	Lucia.Segovia-De-La-Revilla@lshtm.ac.uk
Date of last edit	29/10/2024, version 2.4.0

Guidance on writing a Data Management Plan can be found at
https://lshtm.sharepoint.com/Research/Research-data-management/
and http://servicedesk.lshtm.ac.uk

Advice and feedback can be obtained from:
researchdatamanagement@lshtm.ac.uk

DESCRIBE YOUR RESEARCH

1. What digital resources – data, code, collection tools, etc. - will you collect/obtain and use?

We provide an overview of the digital resources that will be employed in this project. A full description of the datasets that will be used are recorded in the Data Linkage Table ("00_DMP_data-linkage-table.csv") and it will be regularly updated.

Data:

1. Data on soil and crop mineral composition and geographic location collected by GeoNutrition project, it has been published (Gashu et al. 2021; Kumssa et al. 2022) and data is publicly available in Figshare. Data is accessible in comma delimited file (.csv) and excel-spreadsheet (.xlsx) format.
2. Data on soil and crop mineral composition publicly available ((Chilimba et al. 2011), the data was available in (.xlsx) format.
3. Data on biomarkers of micronutrient status (i.e., plasma Se in Women of Reproductive Age), HIV status, geographic location and sociodemographic information collected by national statistical offices in Malawi as part of Demographics and Household Surveys (DHS) will be obtained. Data is accessible in Stata dataset (.dta), comma delimited files (.csv) and shape files (.shp) format.
4. Data on food consumption, geographic location, sociodemographic characteristic, and other variables collected for the Household Consumption and Expenditure Survey in Malawi will be obtained from LSMS/World Bank. Data will be accessed in Stata dataset (.dta) and comma delimited files (.csv) format.
5. Publicly available data on food composition collected by multiple authors and obtained from FAO/INFOODS website. Data is accessible in excel (.xlsx), comma delimited files (.csv) and pdf format.
6. Publicly available data on food and agriculture supply for sub-Saharan region will be obtained from FAOSTAT (Food and Agriculture Organization of the United Nations). Data files will be accessed in comma delimited files (.csv) format.
7. Publicly available data on population and geographic data obtained from WorldPop. Data is accessible in comma delimited files (.csv) format.

Code:

8. Code will be used from the MAPS project and accessed through GitHub. Code is accessible as R scripts and Quarto.
9. Code will be generated and shared through GitHub. Code will be share as R scripts and Quarto format.

2. What hardware and software will be used in your research?

- Hardware: Laptop Dell Latitude 5310 for all the data collection and analysis
- Software: RStudio for data management (recording, storing, reporting, etc.) and for data analysis, MS Excel for data recording and storing, QGIS for spatial analysis, Git and GitHub for version control and data sharing, Zotero and Mendeley for citation management, OneDrive and SharePoint for data storage and data back-up, MS Word for reporting.

3. What data-related activities will be performed during the research?

Task	Description
Data collection	Dataset will be obtained and stored in month 6-12.
Data cleaning	Dataset will be cleaned and shaped in a format that can be used for the research project in month 12-18.
Data analysis	Models will be developed and tested for the research hypothesis and research objectives in month 14-30.
Data sharing	Cleaned datasets and scripts will be shared in GitHub regularly.

4. What quality checks will you perform to ensure resources are fit for purpose?

1. Data collection:
    - Before: All the dataset needed are identified and recorded in the Data Linkage Table.
    - During: Because we will use secondary data, we will collect information on the quality check performed by the data providers, if available.
    - After: Checking that one copy of each dataset and its metadata is stored in the raw-data folder and information is completed in the Data Linkage Table.
2. Data cleaning:
    - Before: Checking that one R-project has been created for the analysis, including the README file. Check that all the data has been collected and that there is one raw copy securely stored and one working copy in the R project for cleaning and analysis.
    - During: Once the data is accessed, checking that the number of observations and variables are the same as reported in the original source. Then, datasets will be cleaned, checked, and verified to find any possible inaccuracies, outliers, missing values, etc. in the original dataset. We will create a protocol for data checking and validation for each dataset. Then, we will share the code for accessing, cleaning, and checking steps in GitHub and protocols will be share through OSF and/or LSHTM Compass to ensure transparency and reproducibility. Backups will be performed using Git for version control.
    - After: Ensure that the code run smoothly and that all the steps are annotated.
3. Data analysis:
    - Before: A statistical plan will be created in MS Word that will be reviewed by the supervisory team and co-authors and that will be share with the Ethics Committee. After approval, the document will be publicly available through OSF and LSHTM Data Compass.
    - During: Keeping a log of the analysis performed and ensuring that the version control is maintained and commits provide useful information.
    - After: Checking that the scripts with the code for the analysis are reproducible and results are consistent.
4. Data sharing:
    - Before: Checking "Term of Use" of each dataset to ensure that we are compliant with any information shared.
    - During: Ensure that reports and code is free of errors, bugs and run correctly. Ensure that citation and data source acknowledgement are correct.
    - After: Keep a record of data and code shared and ensure that it is maintain and up to date until the end of the project.

5. How will you address ethical & legal issues within your research?

Ethics:

- Ethical approval will be requested from the LSHTM Observational Research Ethics Committee.

Data permission:

- Most of the dataset that will be used are publicly available, hence only source acknowledgement and proper citation is necessary for legal compliancy (Resource 1, 3, 4 and 6).
- Household Consumption and Expenditure Survey or biomarkers of dietary status surveys are available for research purposes, we will need to request online access from the WorldBank and sign a Term of Use with DHS program (Resource 2 and 5).
- For biomarker Se data, approval will be sought from Malawi Government Department of Nutrition, HIV/AIDS, Ministry of Health, via the data owner Dr Felix Phiri. A Data Transfer Agreement will be used (Resource 7).
- In addition, we will keep records of each dataset/ organization Term of Use in the Data Linkage Table and will be followed throughout the project.

6. What documentation will be created to ensure resources can be understood?

The project will be documented using R scripts for data management and Rmarkdown for reporting. We aim to make the project fully transparent and reproducible following the principles of Open Science.

STORAGE AND SECURITY

7. Where will resources be stored at key stages of your research?

Data classification:

Data will be classified according to the LSHTM Data Classification and Handling Policy before selecting the most appropriate storage location. Information on the classification will be stored in the Data Linkage Tables for each dataset.

- Confidential: Although Se biomarker data is anonymized, we will treat Se biomarker data as confidential (resource 7).
- Internal: Data on Household Surveys from WorldBank and DHS Program will be treated as internal (resources 2 and 5)
- Public: Data published and available through CC-by dataset (Resource 1, 3, 4 and 6) in addition to the code (Resources 8 and 9) are currently publicly available.

Security and storage location:

- All data used in the project is anonymized prior to access, and we will not attempt to de-anonymize the data. When data transfer will be needed encrypted email and/or OneDrive will be used. We will use 7-Zip for file encryption and password will not be shared along with the encrypted files as recommended by LSHTM.
- Confidential data will be stored in password-protected folders on SharePoint and OneDrive will be used for storing internal and public data.
- Confidential raw data will be only accessed by the student (LSdlR) and the team members included in the Data Transfer Agreement, similarly for internal data only team members under the Term of Usage (ToU) of the respective data owners will be allowed access.
- Public data and summarized data will be made publicly available through publications arising from the work.

8. What labelling conventions will you apply to manage your resources?

Labelling and name conventions will be used following the conventions recommended in [here]() and [here]() (E.g., using ISO 8601 for date format (YYYY-MM-DD). A README will be generated for the whole project with detailed information on the data structure and name convention will be provided and regularly updated.

For data analysis, we will create one R project for each PhD objective that will contain all the information to reproduce the analysis and it will run as stand-alone project. This R project will contain (i) scripts for accessing, cleaning, analysis and visualizing the data, (ii) a data folder with data needed that will not be shared, (iii) a folder with outputs such as reports, tables, graphs, etc. and, (iv) a README file explaining the steps and analysis performed.

For files that are not generated as R project (i.e., scoping review data screening), folders and file names will be organized according to the PhD objective that it is related to, for example: "01" for the scoping review, or "02" for geospatial modelling. The number or identifier will be followed by a short name and/or description of the file: E.g., "01_protocol_food-composition-data-ssa". In addition, we may add version number if the file will be reviewed by other members of the team (i.e., documents shared with the supervisory team for feedback ("01_protocol_food-composition-data-ssa_v1.0")). Finally, files generated from LSHTM templates will be labelled as "LSHTM_" and the file (e.g., "LSHTM_SegoviadelaRevillaL_DMP_v1.1.doc")

9. How will you keep data safe and secure? (choose one or more)

Only anonymised data will be used - personal, sensitive, or otherwise confidential data is not needed for the research	x	Store personal details in a separate secure location & link it via an identifier		Delete personal & confidential details at earliest opportunity (specify when below)	
Use digital storage that require a username/password or other security feature	x	Physical security (such as locked cabinet or room)		Protect portable devices using security features, e.g. biometric	x
Encrypt storage devices		Encrypt during transfer	x	Avoid cloud services located outside EU	x
Take 'Information Security Awareness training'	x	Ensure backups are also held securely	x		
Notes:					
Identify additional steps you will take to avoid, reduce, or eliminate risks that may affect your resources.					

ARCHIVING & SHARING

10. What resources should be kept as evidence of your research?

All resources will be kept for 10 years following completion of the analysis, in line with institutional research data management policy

11. Where will these resources be hosted?

Original dataset will be kept in a LSHTM Secure Server for record purposes, additionally, outputs generated that do not breach the ToU of the original data will be shared in OSF/Zenodo, as well as will share every input needed for replication in GitHub.

12. When will the resources be made available? (choose one or more)

During the research life	x	At the same time as findings are published in an academic journal	x	A set time after research end, e.g. 12 months. Specify below	
Resources already available (provide details below)	x	On completion of my thesis		Other (provide details below)	
Further information / Other					

13. How will you make other researchers aware that the resources exist?

Publish a metadata record describing the resources in a repository or other catalogue	x	Obtain a Digital Object Identifier (DOI) or other permanent ID	x
Cite resources in future research papers, e.g. in the data access statement or reference list	x	Cite resources in project reports	
Publish a description for the project website		Write and publish a Data Paper	
Add resources to a list of your academic outputs			
Other measures / Further details			

14. What steps will you take to ensure resources are easy to analyse and use in future research? (choose one or more)

Prepare a codebook or other documentation that provides an accurate description of content	x	Store resources in open file formats such as CSV, Rich Text, etc. See https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats	x
Write a user guide that provides a high-level overview of research		Apply a standard licence that allows a broad range of uses (e.g. Creative Commons, Open Data Commons)	x
Designate a corresponding author / data custodian who will handle data-related questions	x	Use domain-specific standards that make it easy to import and analyse data	
Other / Further information			

15. If resources can be made available, but not openly, what conditions on access/use must be met?

Requirement:	To be addressed by:
LSMS data from the WorldBank requires registering and agreeing with the Term of Use. The data will not be shared but summaries and aggregated data could be shared and/or published allowing proper citation of the data. More details about Term of Use are described here under (Public Use): https://microdata.worldbank.org/index.php/terms-of-use	Need to create an account and agree with the Terms of Use. Then citation is required when publishing summaries or any other publication.
DHS data requires to be a registered user and request access to the dataset. In addition, for GPS and Biomarkers data a Terms of Use Statement needs to be electronically signed before receiving access. Any report or publication generated from that data must be sent to The DHS Program Data Archive. More details about Term of Use are described in here: (https://dhsprogram.com/data/Terms-of-Use.cfm)	Need to create an account, request access, and sign a Terms of Use Statement. Full details on the access are provided in here: (https://dhsprogram.com/data/Access-Instructions.cfm)

RESOURCING

16. What are the primary data management challenges in your research?

The main challenge is to efficiently manage the multiple dataset that will be used in this project since each dataset has different nature and unique data providers. Hence, keeping a good record of each dataset, metadata, quality etc. is paramount for a good research practice.

17. How can LSHTM & others help you to better manage your data?

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Ms Lucia Segovia de la Revilla
LSHTM

9 May 2022

Dear Ms Segovia de la Revilla,

Study Title: Geospatial analysis of food composition data to estimate population micronutrient intake

LSHTM Ethics Ref: 26546

Thank you for your application for the above research project which has now been considered by the *Observational* Committee via Chair's Action.

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, 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 is as follows:

Document Type	File Name	Date	Version
Other	Research_Ethics_online_training_certificate		
Investigator CV	C_Dooley_CV_28092021	28/09/2021	1
Investigator CV	CV_Edward-Joy-October-2021	31/10/2021	1
Investigator CV	CV_Segovia-de-la-Revilla-Lucia_2022	13/04/2022	1
Protocol / Proposal	LSHTM_Segovia-de-la-Revilla-L_Protocol_v3.1	21/04/2022	3.1

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 the 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.

Further information is available at: www.lshtm.ac.uk/ethics.

Yours sincerely,

[redacted]

ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

This is to certify that

Lucia Segovia de la Revilla

successfully completed the

Research Ethics

e-learning course

with a score of

85.00 %

Comprising of modules covering:

- Introduction to the History of Research Ethics
- Fundamental Ethical Principles, including:
    - Respect for persons
    - Beneficence
    - Justice
- Responsibilities of Research Ethics Committees
- Understanding Vulnerability
- Privacy and Confidentiality

On

July 5, 2021

Provided by

London School of Hygiene & Tropical Medicine

This course meets the requirements for protection of human subjects training required by individuals involved in the design and/or conduct of National Institutes of Health (NIH) funded human subjects research.

List of peer-reviewed publication

1. **Segovia de la Revilla, L**., Codd, T., Joy, E.J.M., Mlambo, L., Grande, F., Rittenschober, D., Moltedo, A., Holmes, B.A., Ander, E.L., (2025). An open science framework and tools to create reproducible food composition data for use in nutrition. *Journal of Food Composition and Analysis* 137, 106894. https://doi.org/10.1016/j.jfca.2024.106894.
2. Goto, R., Mlambo, L., **Segovia de la Revilla, L.,** Swai, A., Mshida, H., Amos, A., ... & Joy, E. J. (2024). Estimating food consumption, micronutrient intake and the contribution of large-scale food fortification to micronutrient adequacy in Tanzania. *Public Health Nutrition*, *27*(1), e230.
3. **Segovia de la Revilla, L.,** Ferguson, E. L., Dooley, C., Osman, G., Ander, E. L., & Joy, E. J. M. (2023). The availability and geographic location of open-source food composition data used to estimate micronutrient intakes in sub-Saharan Africa: A scoping review. *Journal of food composition and analysis*, 120, 105322. doi:10.1016/j.jfca.2023.105322
4. Shewry, P. R., Joy, E. J. M., **De La Revilla, L. S**., Hansen, A., Brennan, J., & Lovegrove, A. (2023). Increasing fibre in white flour and bread: Implications for health and processing. *Nutrition bulletin*, 48(4), 587-593. doi:10.1111/nbu.12648
5. Andrés-Hernández, L., Blumberg, K., Walls, R.L., Dooley, D., Mauleon, R., Lange, M., Weber, M., Chan, L., Malik, A., Møller, A., Ireland, J., **Segovia, L**., Zhang, X., Burton-Freeman, B., Magelli, P., Schriever, A., Forester, S.M., Liu, L., King, G.J., (2022). Establishing a Common Nutritional Vocabulary - From Food Production to Diet. *Frontiers in Nutrition* 9. https://doi.org/10.3389/fnut.2022.928837.
6. Ramos, S., **Segovia, L.**, Melado-Herreros, A., Cidad, M., Zufía, J., Vranken, L., & Matthys, C. (2022). Enviroscore: normalization, weighting, and categorization algorithm to evaluate the relative environmental impact of food and drink products. *NPJ science of food*, 6(1), 54. doi:10.1038/s41538-022-00165-z
7. Sandalinas, F., Filteau, S., Joy, E. J. M., **Segovia de la Revilla, L.,** MacDougall, A., & Hopkins, H. (2022). Measuring the impact of malaria infection on indicators of iron and vitamin A status: a systematic literature review and meta-analysis. *The British journal of nutrition*, 129(1), 87-103. doi:10.1017/S0007114522000757
8. De Bauw, M., **De La Revilla, L. S.,** Poppe, V., Matthys, C., & Vranken, L. (2022). Digital nudges to stimulate healthy and pro-environmental food choices in E-groceries. *Appetite*, 172, 105971. doi:10.1016/j.appet.2022.105971
9. Joy, E. J. M., Kalimbira, A. A., Sturgess, J., Banda, L., Chiutsi-Phiri, G., Manase, H., Gondwe J, Ferguson E.L., Kalumikiza Z, Bailey EH, Young SD, Matandika L, Mfutso-Bengo J, Millar K, NIKSIC M, **Segovia De La Revilla L.,** Likoswe BH, Phuka JC, Phiri FP, Lark RM, Gashu D, Langley-Evans SC, Ander EL, Lowe NM, Dangour A.D., Nalivata PC, Broadley MR. (2022). Biofortified Maize Improves Selenium Status of Women and Children in a Rural Community in Malawi: Results of the Addressing Hidden Hunger With Agronomy Randomized Controlled Trial.. *Frontiers in Nutrition*, 8, 788096. doi:10.3389/fnut.2021.788096
10. Tang, K., Adams, K. P., Ferguson, E. L., Woldt, M., Kalimbira, A. A., Likoswe, B., Yourkavitch J, Chrisinger B, Pedersen S, **Segovia De La Revilla L.**, Dary O, Ander E. L., Joy, E. J. M. (2021). Modeling food fortification contributions to micronutrient requirements in Malawi using Household Consumption and Expenditure Surveys. *Annals of the New York Academy of Sciences*, 1508(1), 105-122. doi:10.1111/nyas.14697

List of pre-print publication:

1. **Segovia de la Revilla, L**., Ferguson, E. L., Dooley, C., Osman, G., & Joy, E. (2022). Assessing the availability and geographic location of food composition data used to estimate micronutrient intake in sub-Saharan Africa: protocol for a scoping review. doi:10.31219/osf.io/vd2mf

2. Ramos, S., **Segovia de la Revilla, L.**, Herreros, A. M., Cidad, M., Vranken, L., & Matthys, C. (2020). Food Enviroscore - Methodology description. doi:10.31219/osf.io/t2hz4

Books and Tutorials

Mlambo L., **SEGOVIA DE LA REVILLA L.,** Codd T., Osman G., Tang, K., Joy E., Ander L.E. (2023) R for Household Consumption and Expenditure Surveys. https://www.maps.africa/r4hces/

Data and Code Sharing (Open Access Programming Software & Code lists)

1. **SEGOVIA DE LA REVILLA L** (2023). LuciaSegovia/scoping-review-fct-ssa: Published version of the code (v.1.0.0). Zenodo. https://doi.org/10.5281/zenodo.10213390
2. **SEGOVIA DE LA REVILLA L.,** Codd T., (2023) Cleaning and standardization of food composition tables and databases for MAPS. Repository. https://github.com/LuciaSegovia/fct
3. Codd T., **SEGOVIA DE LA REVILLA L** (2023). NutritionTools: Tools for the Organisation, Matching, Calculation, and Summarisation of Nutrition Data. R package version 1.0.1, https://tomcodd.github.io/NutritionTools/, https://github.com/TomCodd/NutritionTools
4. GOTO R., **SEGOVIA DE LA REVILLA L** (2023). Household Consumption and Expenditure Survey Analysis in Tanzania National Panel Survey 2014-15 Wave 4. Repository. https://github.com/rgoto55/TNPSW4/tree/main

Conference Presentation (Oral)

1. **Segovia de la Revilla L.,** Dooley, C., Ferguson, E.L., Joy, E.J.M. 2024. Plasma selenium concentrations are associated with spatial variation in maize selenium concentrations in Malawi; implications for assessing dietary selenium adequacy. GEOMED 2024. Hasselt, Belgium.
2. **Segovia de la Revilla L.,** Codd T., Mlambo L., Grande F., Moltedo A., Ander L., Holmes B. 2023. Food composition data: A transparent open framework to compile reproducible food composition tables and databases. 8th ANH Academy Week. Lilongwe, Malawi. Hybrid (in-person and online).
3. Joy E., Ander E.L., Moltedo A., Kiøsterud E., Mlambo L., **Segovia de la Revilla L.,** Osman G. 2023. Side event: New resources for nutrition analysis using Household Consumption and Expenditure Survey data. 8th ANH Academy Week. Lilongwe, Malawi. Hybrid (in-person and online).
4. **Segovia de la Revilla, L.** 2022. What is the food composition data used to estimate dietary intake of calcium, iodine, iron, selenium, and zinc in sub-Saharan Africa? - Results of a scoping review. 7th ANH Academy Week. Online.
5. **Segovia de la Revilla, L**. 2021. Current availability of food composition data in sub-Saharan Africa. 2nd Africa Food Environment Research Network (FERN2021) Meeting & INFORMAS eSymposium Series for Africa. 3rd, 10th & 17th November 2021. Online.
6. **Segovia L.,** Ramos S., Cidad M., Poppe V., Vranken L., Matthys C. 2019. Nutritional quality and environmental footprint of food products – impact of an integrated single score. Belgian Nutrition Society, Ninth annual meeting. Brussels, Belgium.

Conference Presentations (Poster Presentations)

1. **Segovia de la Revilla L.,** Ferguson E., Dooley C., Joy E., 2023. Using geo-referenced maize selenium concentration to estimate plasma selenium concentration in Malawi: a preliminary study. Micronutrient Forum 6th Global Conference. The Hague, Netherlands.
2. **Segovia de la Revilla L.,** Codd T., Mlambo L., Grande F., Moltedo A., Ander L., Holmes B. 2023. Food composition data: A transparent open framework to compile reproducible food composition tables and databases. 8th ANH Academy Week. Lilongwe, Malawi. Hybrid (in-person and online).

3. **Segovia de la Revilla** L., Codd T., Mlambo L., Grande F., Moltedo A., Holmes B., Ander L. 2022. A path to "FAIR" Food Composition Data. 22nd IUNS-ICN International Congress of Nutrition. Tokyo, Japan.
4. **Segovia de la Revilla L.,** Ferguson E., Dooley C., Joy E., 2022. A scoping review of the availability of geospatial information in food composition databases in sub-Saharan Africa. 22nd IUNS-ICN International Congress of Nutrition. Tokyo, Japan.
5. **Segovia L,** Ramos S., Cidad M., Melado-Herreros A., Braems A., Van Hemelryck S., Poppe V., Vranken L., Matthys C. 2019. Is the PEF-based single score a more relevant marker of the environmental footprint of the European food basket?. EAAE Symposium. Montpellier, France.