Tufts UNIVERSITY

Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

13th INTERNATIONAL FOOD DATA CONFERENCE

FEED the FUTURE The U.S. Government's Global Hunger & Food Security Initiative

Innovation Lab for Nutrition

USAID FROM THE AMERICAN PEOPLE

CANDASA

Validation and compilation of regional food composition data for nutrition research: Pilot results for Southeastern Africa

Yan Bai[1], Elena N. Naumova[1], Lynne Ausman[1], Shibani Ghosh1, Averalda van Graan[2], Agnes Mwangwela[3] and William A. Masters[1*]

1 Friedman School of Nutrition, Tufts University. USA. 2 SAFOODS, Biostatistics Unit. South African Medical Research Council. South Africa.
3 Faculty of Food & Human Sciences. Lilongwe Univ. of Agric. and Natural Resources. Malawi

Background and Aims

- Reliable food composition data are essential for nutrition and public health research and education, foods development and trading, as well as nutrition, health and agriculture policy making;
- Significant variation in nutrient composition of foods results from gene-environment interaction, modified by storage and processing. Key influences include feed, soil, climate, genetic resources (varieties/cultivars, breeds), storage conditions, processing, fortification and market share[1];
- We introduce a method to identify the biologically plausible range of variation across a geographic region, combining food composition tables (FCT) with statistical tests to distinguish biologically plausible variation from non-classical measurement error;
- Our aim is to identify the biological variation in nutrient composition that might be associated with genetic diversity, interacting with differences in agronomic and ecological conditions across countries;
- With sufficient observations, the statistical distribution of many variables, such as plant height or leaf size as well as nutrient composition, tends to follow a right-skewed distribution with extreme values supported by other extreme values in a log-normal distribution[2,]

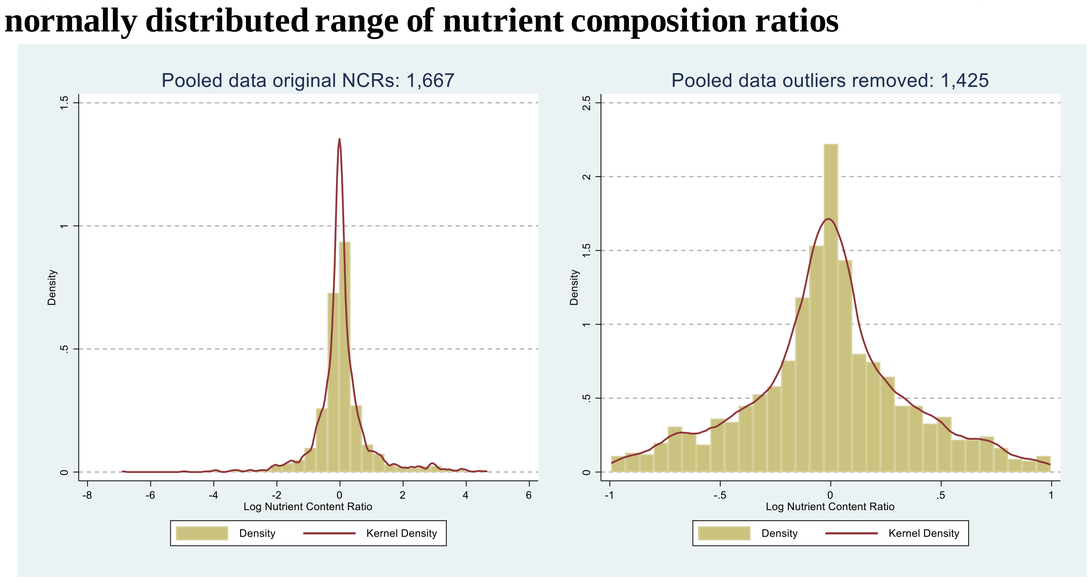
Methods and Data

- To test across diverse nutrients we compare unit-free nutrient content ratios (NCR) for each nutrient in each food to a common benchmark, which for maximum sample size is the SR28 dataset from USDA (2016), and then flag as implausible the smallest set of observations needed for the remaining data to be drawn from a lognormal distribution and therefore likely to reflect biological variability;
- Food items considered for this study include only standardized items for which market prices are collected by national statistical agencies, totaling 55 foods in Malawi and 71 in Tanzania;
- To construct a combined FCT covering as many items as possible, we merge their nutrient profiles from the Malawi and Tanzania FCTs with USDA standard reference (SR28) data. Items may have nutrient composition data from one, two or all three FCTs. After matching item descriptions, confirm that nutrient profiles refer to the same kind of food by comparing moisture content and energy density;
- We matched 44 and 51 food items in Malawi and Tanzania respectively to the SR28 benchmark, for a total of 61 unique items in the regional dataset, leading to a universe of 770 values for 31 compositions in Malawi, and 897 values for 30 compositions in Tanzania and a pooled set of 1,667 NCRs.

Results

- From the pooled set of 1,667 NCRs, removal of 242 extreme values leaves a regional dataset for which log-normality cannot be rejected (p=0.0517);
- Values in this regional FCT range from 0.37 to 2.7 times their counterparts in SR28;

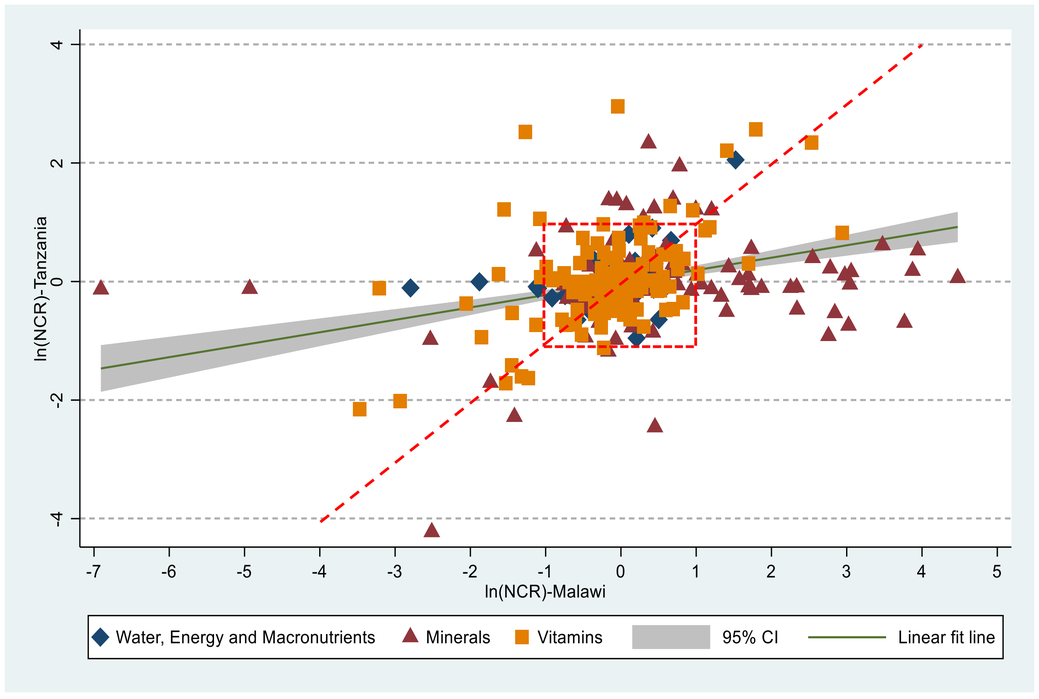
Fig 1. Omission of extreme values leaves a biologically plausible, log-normally distributed range of nutrient composition ratios



Note: Data shown are histograms and estimated distributions of log-transformd ratios of each nutrient/food density to a common benchmark set of values (USDA SR28). The Stata sktest for skewness is used for the test of log-normality.

- Implausible observations are most common among minerals and least likely for macronutrients, which could guide further investigation.

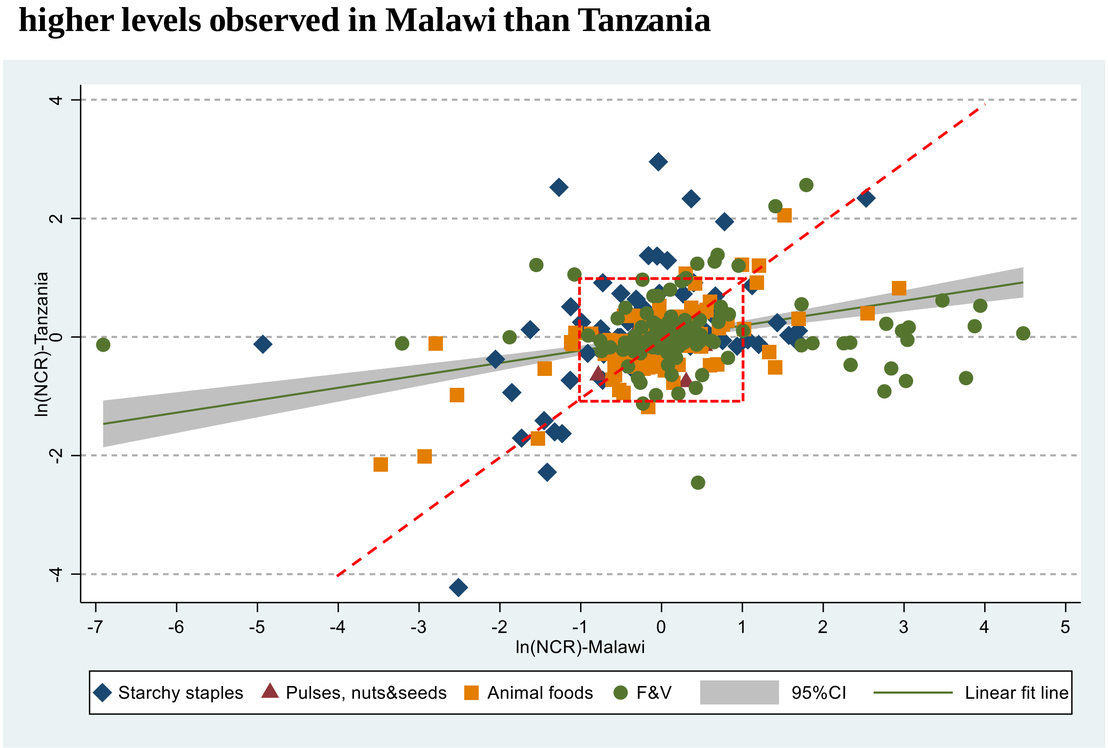
Fig 2: The most common extreme values are for minerals, with higher levels observed in Malawi than Tanzania



Note: Data shown are scatter plots of log-transformed ratios for each nutrient/food density, relative to a common benchmark set of values (USDA SR28), for the 346 nutrient-food pairs observed in both Malawi and Tanzania. A 45-degree line would indicate identical values observed in the two countries, and proximity to zero indicates similarity to SR28.

- Similarly, we also observed most implausible values among fruits and vegetables while least among pulses, nuts and seeds;
- Although extreme values exits in starchy staples and animal foods, the deviations in both countries are more aligned.

Fig 3: The most common extreme values are for fruits and vegetables, with higher levels observed in Malawi than Tanzania



Note: Data shown are scatter plots of log-transformed ratios for each nutrient/food density, relative to a common benchmark set of values (USDA SR28), for the 346 nutrient-food pairs observed in both Malawi and Tanzania. A 45-degree line would indicate identical values observed in the two countries, and proximity to zero indicates similarity to SR28.

Conclusions and Limitations

- The relevance and reliability of locally available FCT data provide fundamental support to nutrition, dietetics, health and food studies in the world;
- High quality FCT data are also critical for developing effective nutrition sensitive agriculture and food policies;
- Our study proposes a robust framework to validate and compare different FCT data sources, which may target the most implausible data in a given dataset for further investigations, and therefore to be applied to improve FCT data quality in developing and low-income countries;
- Our method is readily scalable, using additional observations to expand the plausible range of data, and lead to compile regional FCT data bases, which are highly needed in comparative studies across regions and the globe;
- Future work is needed to understand specific mechanisms leading to data deviations, therefore to further improve the precision of the method.

Acknowledgements

This work was funded by the Bill & Melinda Gates Foundation through the CANDASA project (OPP1182628), for the purpose of constructing price indexes for access to nutritious diets in Malawi, Tanzania and elsewhere. In so doing we make use of a draft food composition table for Malawi constructed with support from the Feed the Future Innovation Lab for Nutrition (AID-OAA-L-10-00006) funded by the United States Agency for International Development.

References

1. Food composition challenges. INFOODS: Food composition challenges Available at: http://www.fao.org/infoods/infoods/food-composition-challenges/en/. (Accessed: 2nd August 2019)
2. 1. Berti, P.R. and Jones, A.D., 2013. Biodiversity's contribution to dietary diversity. Chapter 9 (pp. 186-206), in J. Fanzo et al., eds., Diversifying food and diets: using agricultural biodiversity to improve nutrition and health. New York: Routledge.