Data and code for study: "Food-sourcing from on-farm trees mediates positive relationships between tree cover and dietary quality in Malawi"

These files contain data collected from 460 Malawian households in October 2021 (dry season) and March 2022 (wet season). It also contains a list of unique food items collected through the 24-hr dietary recall survey. Using publicly available nutrition data (see below), the nutrient value of each food item/recipe was recorded. This study-specific food composition table (FCT) was used to calculate the micronutrient adequacy of individual respondents (see Methods).

We have also included the code necessary to re-produce the causal mediation analysis (and associated sensitivity analyses) featured in the study (format: R Script).

Description of the data and file structure

The files include:

1. an Excel workbook (Excel 2016) with the following sheets:
    1. FCT: the study-specific food composition table (see above) - to be used to refer to the nutrient value of specific recipes/items consumed by study respondents
    2. FCT key: a key describing the color codes for different values in the FCT
    3. allfooditems: a list of unique food items consumed by study respondents
    4. Dry Season: dry season data for all variables (covariates and outcomes) included in the models - to be used to reproduce the causal mediation analysis and associated sensitivity analyses/robustness checks
    5. Wet Season: wet season data for all variables (covariates and outcomes) included in the models - to be used to reproduce the causal mediation analysis and associated sensitivity analyses/robustness checks

The following variable codes are used to describe the dry and wet season data. The codes can also be used to understand the R Scripts.

Variable codes	
FC.1km	Percentage of tree cover within a 1km radius of the household (Reiner et al., 2023)
FC.500m	Percentage of tree cover within a 500m radius of the household (Reiner et al., 2023)
north.south	Study region
edu.lvl	Level of education obtained by the respondent (note: in categories 'Diploma' and 'Secondary' were collapsed for the analysis)
hh.size	Number of people living in the household
hh.mpi	Multidimensional Poverty Index (MPI) living standard dimension (0-1) (Alkire & Santos, 2014)

no.acres.cult	Number of acres cultivated in the last growing season (2021) by the household
TLU	Livestock holdings, measure in tropical livestock units (FAO, 2018)
crop.count	Number of crops cultivated by the household in the last growing season (2021)
treeuse.food	Whether or not the household sources food from trees on or around (within 15 m) of their farmland
food.trees	The number of unique food tree species cultivated by the household
VitA.MSM.NAR	Nutritional adequacy ratio (0-1) for vitamin A intake relative to recommended intake
Iron.MSM.NAR	Nutritional adequacy ratio (0-1) for iron intake relative to recommended intake
Zinc.MSM.NAR	Nutritional adequacy ratio (0-1) for zinc intake relative to recommended intake
Folate.MSM.NAR	Nutritional adequacy ratio (0-1) for folate intake relative to recommended intake
MAR	Mean Adequacy Ratio: the average NAR across vitamin A, zinc, iron, and folate intakes

References

Alkire, S., & Santos, M. J. (2014). Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index. World Development, 59, 251–274. https://doi.org/10.1016/j.worlddev.2014.01.026

FAO. (2018). Guidelines on methods for estimating livestock production and productivity. Rome.

Reiner, F., Brandt, M., Tong, X., et al. (2023). More than one quarter of Africa's tree cover found outside areas previously classified as forest. Nature Communications (2023). preprint available at https://www.researchsquare.com/article/rs-1816495/v2

Sharing/Access information

The following contains information on the publicly available datasets used in this analysis.

We accessed very high-resolution (3m) tree cover data from 2019 PlanetScope nanosatellite constellation imagery: https://www.nature.com/articles/s41467-023-37880-4

Food consumption data (used to compile the study-specific Food Composition Table (FCT), was sourced from the following publicly available datasets:

Malawi: http://hdl.handle.net/10427/D217R336D

Tanzania: https://nutritionsource.hsph.harvard.edu/food-tables/

Zambia: https://nfnc.org.zm/download/zambia-food-composition-tables-4th-edition/

Mozambique: http://hdl.handle.net/10138/337295

Kenya: https://nutritionhealth.or.ke/programmes/healthy-diets-physical/food-composition-tables/

West Africa: https://openknowledge.fao.org/server/api/core/bitstreams/c5b37ac2-7082-48ab-a4a5-68d27deb4849/content

United States (USDA): https://fdc.nal.usda.gov/

Code/Software

The data to be uploaded in R is included as an Excel file (Excel 2016). Note: if you want to run the code, you will have to save the Excel sheets containing dry season data and wet season data as separate files with names that correspond to the R Script.

To examine the relationship between tree cover, use of on-farm food trees, and dietary quality, we cleaned and analyzed our data in R (version 4.2.2). The code included here includes the linear regressions, controlling for a selection of covariates, that we ran to examine these relationships. Specifically, we used the 'mediation' package in R to evaluate the causal mediation (i.e. indirect) effect of our food tree variable. This package also allowed us to evaluate the robustness of our average causal mediation effect estimates by testing for the possible confounding effect of unmeasured pretreatment variables. We also tested for the impact of omitted variables using the 'sensmakr' package in R.

Other R packages used to prepare the data:

(openxlsx)

(readxl)

(dplyr)

(scales)

(regclass)

(forcats)

The code for the causal mediation analysis and associated sensitivity checks is formatted and uploaded as an R Script file.