MAX PLANCK INSTITUTE
FOR EVOLUTIONARY ANTHROPOLOGY

STUDY INFORMATION AND DATA USE AGREEMENT FOR THE RESEARCH PROJECT: "*CROSS CULTURAL COMPARISON OF HUMAN GROWTH TRAJECTORIES*"

Catalina I. Fernández, Caissa Revilla-Minaya & John A. Bunce

Max Planck Institute for Evolutionary Anthropology,
Deutscher Platz 6, 04103 Leipzig, Germany

1. Brief Description

In this project, we will apply the growth model recently developed by Bunce, Fernández, and Revilla-Minaya 2022 (see: Causal models of human growth and their estimation using temporally-sparse data) to longitudinal datasets from populations inhabiting a wide range of ecoregions with the aim of distinguishing and describing general patterns and global variation in human growth trajectories. By using this model, we expect to be able to estimate, and hopefully disentangle, the relative contributions of different factors affecting growth in height and weight around the world, as well as to compare specific components of growth among diverse populations.

2. Intended use of the Data

The data provided by the collaborators in this project will be used for the sole purpose of this study on the comparison of growth trajectories. The final product will be one or more scientific articles on this topic. Data will not be used for any commercial purposes. We will not distribute these data to third parties under any circumstance. We will make no attempt to identify or contact research participants, their households or communities. Collaborators will have the option of making the de-identified data they contribute publicly available as an accompaniment to the published article(s) that result from this project (see below).

3. Dataset and population (s)

Please provide the following characteristics of the dataset (s), indicating population, location (s), a brief description of data collection methods and/or a reference for them, and indicate whether data was collected after IRB approval or other ethics review (please, indicate review board and protocol number). If other ethical protocols were followed, please specify.

- **Population name(s) and brief description (e.g., subsistence practices, refugees from another region, access to Western healthcare systems, common health challenges, etc.):**

These data are from children in the Machame area of Kilimanjaro, Tanzania. Machame is a rural, agricultural area. By car, it is about one hour's drive to the urban center of Moshi (~225,000 population) and two hour's drive to the larger urban center of Arusha (~500,000). Most residents of Machame are members of the Chagga ethnic group; Chagga are historically agropastoralists (but not nomadic) and both green bananas and cows remain central to their identity. They have a long history (~1000 years) with the pastoralist Maasai living in the areas surrounding Kilimanjaro. Most long-term residents of Machame speak Swahili and Chagga. While most Chagga throughout Kilimanjaro are Christian (multiple sects), ~25% of Chagga in Machame are Muslim.

The volcanic soils of Kilimanjaro are quite productive and support a fairly dense population all the way to the boundary of Mt. Kilimanjaro National Park. Most households engage in agriculture and grow a mix of subsistence (e.g., cooking bananas) and cash (e.g., coffee, maize) agriculture. It is very common for households to keep cattle (or goats), as well, for both household consumption of the milk and eventual sale for meat. Kilimanjaro is a fairly prosperous region of Tanzania and nutrition is better than in many other places, although some families live in poverty and there are lean times of year for households reliant primarily on their own food production (banana fruit production is low by the end of the dry season).

Piped water for drinking has become increasingly available over the last 30 years, but infrastructure for sanitation is lacking; virtually all households use pit latrines. The high population density combined with limited septic infrastructure means that rates of infectious disease are generally high. Malaria has declined in the last decade as a public health concern (a combination of effective control efforts and better diagnostics), but it is still present, and respiratory infections, diarrheal diseases, soil transmitted helminths, and skin infections are all common.

People have access to affordable, basic medical care in the village; travel can be cost-prohibitive for those who need to access more intensive care. Participation in pre-natal care and MCH care programs has been quite high for as long as I have been working in Machame. As soon as a large rural hospital further up the mountain made hospital delivery care available without charge, hospital birth became almost universal. Parents are generally eager to access education for their children, but school fees after primary school remain cost prohibitive for many families.

- **Location (s) (please be as specific as possible, and include GPS coordinates if available):**

  -3.2826726898646563, 37.2247944804334

  (Lat and long for the clinic that hosted our data collection; let me know if these don't work.)

- **Data collection methods (please include reference if available):**

March 2021-June 2023

Children visited ~monthly for one year

Height was estimated with a Seca portable stadiometer (Seca 213)

Weight was estimated with a digital scale (Seca 803)

Study staff performing anthropometry were healthcare providers (mostly nurses or nurse's aides) who received virtual training from me and hands-on training from team members who had been trained by me in person prior to 2020. Single measurements were taken for both height and weight. I will upload our anthropometry protocol along with the data.

Dates of birth we collected from clinic records (when available) or were reported by parents.

- **IRB/ethical review board approval (yes/no, explain):**

The Kilimanjaro Christian Medical University College Research Ethics and Review Committee (CRERC) and the National Institute for Medical Research both reviewed and approved the study. Permission to conduct the study was granted by the Regional Medical Officer of Kilimanjaro and the District Medical Officer of Hai. Binghamton University's IRB relied on the review and oversight of the CRERC.

4. Full name, affiliation and e-mail of all contributors/collaborators for this dataset (s):

Katherine Wander
Binghamton University (SUNY) Department of Anthropology
katherinewander@binghamton.edu

Blandina Mmbaga
Kilimanjaro Clinical Research Institute
Kilimanjaro Christian Medical Centre
Duke Global Health Institute
b.mmbaga@kcri.ac.tz

Ireen Kiwelu
Kilimanjaro Clinical Research Institute
Kilimanjaro Christian Medical Centre
i.kiwelu@kcri.ac.tz

Brenda Kitali
Kilimanjaro Clinical Research Institute
Kilimanjaro Christian Medical Centre
brendakitali@gmail.com

Megan Gauck
Binghamton University (SUNY) Department of Anthropology
mgauck1@binghamton.edu

5. Data submission

Please send a **link to a folder** containing a file or several files (in case of more than one dataset) in **CSV format** to catalina_fernandez@eva.mpg.de.

- Please include the **name of the population in the file name**, and an additional dataset identifier in case of multiple datasets.
- If the dataset also contains single (as opposed to longitudinal) measures for some individuals, please do not remove them from the file. We will also use those observations in the analysis.
- If you have data on adult individuals, whether or not they were also measured as children/adolescents, please include these observations as part of your dataset.

Please, do not include any direct identifying information in the dataset relating to the individuals, their relatives or household members, such as names, addresses, telephone numbers, e-mail addresses or social media identifiers, etc.

For each dataset, we request collaborators on this project to share a **deidentified** data file (.csv) containing the following information and columns:

Column 1: Individual identification code or number that is consistent across longitudinal observations of the same individual.
Column 2: Date of birth in **dd/mm/yyyy** format.
Column 3: Date of data collection event in **dd/mm/yyyy** format.

* If date of birth is not available, please provide **age** in days, whenever possible, for each data collection event.
* If known age is uncertain, if possible, please provide a **range** of minimum, estimated, and maximum age, based on your best approximations.

Column 4: Sex; coded as **m** or **f**.
Column 4: Height in cm
Column 5: Weight in kg

*Note that, for longitudinal measures, you will have multiple rows with the same individual identification code, date of birth, and sex, but with different dates of data collection, heights, and weights.

6. Long term archiving and data access

One of the objectives of this project is to make the deidentified datasets and code (s) used in the data cleaning and analyses available in an open-access repository. We encourage all collaborators to commit to share and deposit deidentified data in a public repository created for this project, once they have checked that, by doing so, they are not violating compliance with their ethical review protocols (e.g., IRB) or other agreements with research participants and community members/leaders. Making data open-access is <u>not a requirement</u> to participate in this study, but we believe that by granting public access to the code and data used we will allow other researchers to check our analyses and reproduce our findings.
If you agree that data from your research site can be made publicly available for this purpose, we will create an appropriate repository (e.g., on Github) to curate this dataset and indicate the contact information for the researchers responsible for each field site, in case someone wishes to use the dataset for a purpose other than simply checking the results of our analysis. If for any reason, in the future, you wish to change your decision regarding data access and usage, we can either include or remove the data from this repository.

Use of the data and access on a public repository (choose <u>one</u> option):

___X___ Restricted use. I agree that deidentified data can be made publicly available for the sole purpose of reproducing this specific analysis. Data cannot be used by the scientific community or the public for any other purpose.

_____ Unrestricted. I agree that deidentified data can be made publicly available without barriers to access or use.

_____ I do not consent to making this data available in a public repository. I am sharing this data only for the purpose of this study. I do not agree that this dataset(s) or any portion of it/them can be shared publicly under any circumstances.

7. Data sharing information to the communities involved

We believe that one of our responsibilities as scientists is to communicate research findings with the communities and individuals who provided their time and biometric information in order to make this study possible. We are aware that most, if not all, collaborators on this project have active field sites and engage regularly with the communities they work with for the purpose of informing and sharing research findings and other initiatives alike. We ask collaborators to this project to share the results of this study with the contributing communities, particularly the parts that concern the specific population that they work with. If a collaborator is no longer in contact with the study population, please let the project leaders know so that together we can potentially brainstorm an alternative solution. Ideally, presenting results to participants will occur prior to publication, so that participating communities have the opportunity to (re)express their permission for us to publish the results. Sharing the results may take the form of a live Power Point presentation, a video, a written information sheet, or any other format that the collaborator believes is most culturally appropriate to present this kind of information to the communities and participants who provided the data. If it is of interest as a model or guide, the project leaders (Catalina Fernández, Caissa Revilla-Minaya and John Bunce) can share with the contributors the materials that they will design for the Matsigenka population that they work with.

8. Manuscript authorship

The project leaders (Catalina Fernández, Caissa Revilla-Minaya and John Bunce) will draft the main manuscript and decide the order of authorship and co-authorship. We will ask all collaborators for feedback once analyses are completed and while we draft the discussion. We are currently considering organizing a workshop at MPI in Leipzig for project collaborators, after analyses are completed, but prior to publication, in order to share the results of the analyses, discuss the main findings and potential causal mechanisms contributing to variation in growth in different populations. This can also be an opportunity for planning further analyses and related projects for the future. More information about this potential meeting/workshop will be shared in the next months.

9. Tentative timeline

- 30 June 2023: All datasets have been submitted to the project leaders.
- 30 September 2023: All datasets have been checked for errors and formatted for analysis.
- 30 March 2024: Data analysis is complete; manuscript is drafted by project leaders.
- Mid-May 2024: Potential workshop for contributors/collaborators at MPI Leipzig.
- 30 July 2024: Results have been presented to participating communities.
- 30 August 2024: Main manuscript is submitted for publication.

***Note that this is a very tentative timeline.**

10. Questions

For questions regarding data submission, data archiving, and other aspects of this study, or additional information, please contact Catalina Fernández by e-mail at catalina_fernandez@eva.mpg.de.

Name of the person in charge of the dataset:
Katherine Wander
Date:
9 Oct 2023