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.):**

Samburu: The studies are with rural Samburu pastoralists in two subregions: highlands (ldonyio) and lowlands (lpurkel). On ldonyio, some families do a little bit of gardening but rainfall is mostly prohibitive. Families raise cattle, goats, and sheep, with some petty hawking and wage labor. In the lpurkel, rainfall is inadequate for gardening. Families raise cattle, goats, sheep, and camels. The population is more dispersed in the lpurkel compared to ldonyio. Health and educational facilities are inadequate throughout Samburu but the situation is worse in the lpurkel. There have been improvements over time and health and education are county priorities.

There will be 2 datasets: IRB Protocol 140527 (anthropometric data collected 2016 & 2017 from two studies under the same protocol) and Protocol 170509 (anthropometric data collected 2018-2019 & 2022). For 140527, some additional longitudinal measurements might be possible by hand matching to a much older dataset (2009), but so far, it looks like too few individuals to make it worthwhile.

All data is based on research approved by the Western Michigan University IRB and Kenya's national research licensing body (NACOSTI). The data predate recent rules that would require research partners to be asked if they approve use of data by other researchers. Collaboration and involvement of the original researcher conforms to the agreement with families. Use for replication by other researchers would require reconsenting.

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

  Specific GPS coordinates at the community level are not permitted under the IRBs for these studies (there can be as few as 400 individuals at that level, which could allow identities to be revealed). The boundaries for lowlands versus highlands can be provided however, and lowland versus highland residence will be indicated in the data. This is important, because of climate differences between these two subregions.

  Catchment for families in lpurkel ('lowlands') of Samburu (it's not a square, it meanders):
  Approximate northeast corner: 1.2575156844954745, 37.67575780260398
  Approximate southeast corner: 0.6491008205413167, 37.714592280291484
  Approximate southwest corner: 0.8663120380411073, 36.981091870368424
  Approximate northwest corner: 1.3168232282497478, 36.90941941779709

Catchment for families on ldonyio ('highlands') of Samburu (also not square):
Approximate northeast corner: 0.996795348667213, 36.82416569743125
Approximate southeast corner: 0.7897932769953444, 36.9779508418868
Approximate southwest corner: 0.842328706032154, 36.69954192753482
Approximate northwest corner: 0.947968410485998, 36.70296849705357

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

Ethnographic participant-observation, structured interviews, and anthropometric measurements were undertaken in both studies.

For IRB Protocol 140527, immune system biomarkers (Epstein-Barr virus antibodies; C-reactive protein), mineral nutrients (measured in dried blood spot), and heavy metals (measured in dried blood spot) were also collected (relevant papers: Mineral nutrition of Samburu adolescents in the context of drought, polygyny, and pastoralism. Lora Iannotti, Carolyn Lesorogol, Charles Hilton, Charles Owuor Olungah, Theodore Zava, Genevieve Neyland, Belinda Needham, Yuhan Cui, Ellie Brindle, Bilinda Straight**.** *American Journal of Physical Anthropology*. Doi: 10.1002/ajpa.24438; Heavy metal blood concentrations in association with sociocultural characteristics, anthropometry and anemia among Kenyan adolescents. J. Ashley-Martin, Lora Iannotti, Carolyn Lesorogol, Charles Hilton, Charles Owuor Olungah, Theodore Zava, Belinda Needham, Yuhan Cui, Ellie Brindle, Bilinda Straight. *International Journal of Environmental Health*. International Journal of Environmental Health Research. https://doi.org/10.1080/09603123.2021.1929871; another paper in review).

For IRB Protocol 170509, DNA methylation, telomere, and oxytocin receptor assays were also performed (relevant papers: Family socioeconomic status and child telomere length among the Samburu of Kenya. Belinda L. Needham, Bilinda Straight, Charles Hilton, Charles Owuor Olungah, Jue Lin. *Social Science & Medicine* 283: 114182. Doi: https://doi.org/10.1016.j.socscimed.2021.114182; Drought, psychosocial stress, and ecogeographical patterning: Tibial growth and body shape in Samburu (Kenyan) pastoralist children. Bilinda Straight, Charles E. Hilton, Amy Naugle, Charles Owuor Olungah, Duy Ngo, Xi Qiao, Belinda L. Needham. *American Journal of Biological Anthropology*, 2022; 1-19. https://doi.org/10.1002/ajpa.24529; Epigenetic mechanisms underlying the association between maternal climate stress and child growth: Characterizing severe drought and its impact on a Kenyan community engaging in a climate change-sensitive livelihood." Bilinda Straight, Xi Qiao, Duy Ngo, Charles E. Hilton, Charles Owuor Olungah, Amy Naugle, Claudia Lalancette, Belinda L. Needham. *Epigenetics*. https://doi.org/10.1080/15592294.2022.2135213; others in review).

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

Yes: Western Michigan University Protocols 140527 and 170509.

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

Bilinda Straight (Bilinda.straight@wmich.edu), Western Michigan University, School of Environment, Geography, and Sustainability; *Principal Investigator* on both data sets

Charles Owuor Olungah (owuorolungah@uonbi.ac.ke), University of Nairobi, Institute of Anthropology, Gender & African Studies; *Senior Personnel* on both datasets

Charles E. Hilton (hiltonch@email.unc.edu), University of North Carolina at Chapel Hill, Department of Anthropology; *Senior personnel* on both datasets; he directly supervised all anthropometric measurements in the field

Belinda Needham (needhamb@umich.edu), University of Michigan School of Public Health, Department of Epidemiology, *Senior Personnel* on 170509, *Collaborator* on 140527

Other collaborators were involved for biomarkers, psychological component, or statistical models, and thus not relevant co-authors for your study.

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 not a requirement 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 one option):

_____ **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.

___X___ 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: Bilinda Straight

Date: October 14, 2023