



**KAMUZU  
UNIVERSITY**  
OF HEALTH SCIENCES



**University of  
Zurich**  
Department of Economics

**Stanford**  
Center on Early Childhood



**CCWD**  
CENTER FOR CHILD WELL-BEING  
& DEVELOPMENT

# USING WEARABLE SENSORS IN PRACTICE

A user guide for collecting data on parental inputs and  
child development using wearable devices







# Table of Contents

---

<b>1. Introduction</b>	<b>4</b>
<b>2. Child development, parental beliefs and child health</b>	<b>6</b>
2.1. Collect survey data	8
2.2. Analyze survey data	9
<b>3. Phone-based survey</b>	<b>11</b>
3.1. Collect phone-based survey	12
3.2. Analyze phone-based surveys	14
<b>4. Anthropometric measurements</b>	<b>16</b>
4.1. Collect anthropometric measurements	17
4.2. Analyze anthropometric measurements	20
<b>5. Language interactions</b>	<b>21</b>
5.1. Collect language interactions data	22
5.2. Analyze language interactions data	25
<b>6. Social interactions</b>	<b>27</b>
6.1. Collect social interactions data	29
<b>7. Socio-emotional development</b>	<b>33</b>
7.1. Collect electrocardiogram data	35
7.2. Analyze electrocardiogram data	38
7.3. Clean outputted data	40
<b>8. Cognitive development</b>	<b>42</b>
8.1. Collect electroencephalogram data	44
8.2. Analyze electroencephalogram data	50
<b>9. Metadata</b>	<b>52</b>
<b>10. Data quality check-lists and must dos</b>	<b>56</b>
<b>11. Appendix A: Legal and ethical considerations</b>	<b>67</b>
<b>12. Appendix B: Vendor TOR template</b>	<b>73</b>

1

---

# Introduction



**Authors:**

Guilherme Lichand and Nicolò Tomaselli

## 1. Introduction

Between 2020 and 2022, a team of researchers developed a research project to employ user-friendly wearable devices to measure child development in Malawi. With the help of experts from several academic institutions, including Kamuzu University of Health Sciences (KUHeS) and the University of Zurich (UZH), several field trials were conducted so that the field team, led by KUHeS, could gain first-hand experience on how to deploy this innovative method of collecting child development measurements, including biomarkers.

This user guide is intended as a practical tool for development practitioners and implementers interested in conducting a similar exercise. Although much of the information in this manual is heavily based on the work done in Malawi, it is hoped that it can be widely used beyond the context of the original study. In practice, this document is operations-specific and describes the activities required to perform tasks in accordance with good practice, regulations, laws, or management standards. Clear guidance and instructions are provided on the steps needed to complete the various data collection processes.

The manual should be used to ensure that data collection with wearable technologies is conducted consistently and efficiently and to reduce the risk of errors or omissions. The user guide applies to anyone responsible for drafting, reviewing, approving, or distributing survey plans to conduct relatively large-scale data collection exercises. It applies to all countries, but is based on the procedure developed in Malawi by KUHeS and UZH. It should be used wherever there is a need to gain new knowledge about child development or to monitor the progress of programs and projects that aim to improve the situation of children and youth in developing countries using evidence and innovative technology.

The person assigned to use this manual should have some experience in research, monitoring and evaluation, and be familiar with basic technology devices and information and communication technologies (ICTs). The procedures outlined in this manual can be used to train enumerators, supervisors, staff, interns, and volunteers and should be followed to provide structure and guidance to ensure consistent implementation within the organization.

The user guide is organized as follows. Chapter 2 outlines the information to be collected on child development using traditional computer assisted personal interviewing (CAPI). Chapter 3 describes procedures for setting up repeated phone-based data collection, while Chapter 4 delves into methods of collecting anthropometric data. Chapter 5 discusses how to measure language interactions through environmental recordings, and Chapter 6 discusses the use of proximity sensors to characterize social interactions. Chapters 7 and 8 describe the use of ECG portable devices and EEG headbands to measure electrocardiogram (ECG) and encephalogram (EEG), respectively. These two measurements provide proxies for social-emotional development and self-regulation on the one hand and cognitive development on the other. Chapter 9 briefly discusses the role of metadata, while Chapter 10 provides data checks and checklist templates that could be used in the field.

# 2

---

Child development,  
parental beliefs and  
child health



**Authors:**

Guilherme Lichand, John Phuka, Pedro Carneiro,  
Nicolò Tomaselli, and Beverly Laher

## 2. Child development, parental beliefs and child health

This chapter presents the type of information that can be collected to measure a child’s physical, cognitive, and social-emotional development. We focus on what information could be collected, what it should be, and why it is useful to measure these indicators. This list is not exhaustive, but it provides a fairly comprehensive overview of the state of the art in generating reported data on child development in developing countries.

Among the most widely used data collection techniques, including for the collection of child development data, is self-reported information that investigators obtain from interviews with the children themselves or, more often, with their guardians. Yet, it is widely accepted that children under a certain age should not receive questionnaires or tests, and it is therefore appropriate to separate the sample of participants according to the age of the child in the family. When the child is over the cut-off age (e.g., 3 years old), the interviewer could administer some of the survey forms to the child, otherwise all forms are administered to the guardian.

The purpose of this kind of data collection is to obtain reference information about the study participants. What is often referred to as the “baseline,” by definition takes place before any other activity. Thus, a face-to-face baseline survey is independent of other data collection activities, and is usually supported by the computer-assisted personal interview (CAPI); software and tablets are used to program and administer the survey instrument.

For the survey to be complete, a multi-module instrument must be administered to all individuals participating in the research. This is essential to ensure that the status of the participants is known, so to study the heterogeneity of the sample and its evolution overtime. This information will be used by practitioners or researchers in combination with data coming from different sources.

There are several validated questionnaires to properly measure children’s cognitive and social-emotional skills. These forms depend on age and context. An internationally validated tool is the International Development and Early Learning Assessment (IDELA), while the Government of Malawi has developed a nationally validated tool called the Malawi Development Assessment Tool (MDAT).

Face-to-face data collection may include the following measures, among others.

Module	Respondent
Household Roster & Characteristics	<i>Primary Respondent / Guardian</i>
Guardian Characteristics	<i>Guardian</i>



Module	Respondent
Child Health	Guardian
Child Functioning	Guardian
Child's Behaviour	Guardian
Guardian's activities with children	Guardian
Parental Attitude	Guardian
Language Interactions	Guardian
Self-Awareness	Child
Child Development	Child
Social Emotional skills	Child
Executive-Functions	Child
Oral Comprehension	Child
Letter Identification	Child

## 2.1 Collect survey data

Data from a survey instrument with numerous modules or sections, administered face-to-face with respondents, are called multi-module survey data. Each module often focuses on a particular topic or idea. When considered as a whole, it creates a comprehensive survey instrument that collects information on a variety of topics.

To collect these data, at least the following steps need to be taken:

1. State the research questions the survey will seek to answer. This will help establish the purpose of the survey and guide the choice of survey forms.
2. Sample selection. Due to budgetary constraints, any data collection must be limited in scope; therefore, it is essential to carefully select the sample of respondents. To be considered valid for statistical analysis, two basic characteristics that a sample must possess are randomness and representativeness. Sample randomness means that every single unit in the population has an equal chance of being selected for the sample. Representativeness means that the subset of the population chosen as the sample accurately reflects the members of the entire population.
3. Create the survey. Design a survey instrument with each module that must answer the study question(s). It is critical to make sure that the survey is clear, concise and easy to interpret.
4. Choose a survey tool. Select a suitable Computer Assisted Personal Interviewing (CAPI) tool to create and administer the survey instrument. At the time this user guide was written, SurveyCTO (<https://www.surveycto.com/>) was a popular survey tool that allows researchers to design and customize their own survey instruments, including multi-module surveys, and distribute them through face-to-face interviews. It also offers features such as skip patterns and validation checks to ensure data quality. Other tools are similarly effective.

5. Train survey staff. Select wisely and train data collection staff on the use of the survey instrument and the proper administration of survey forms. This training should include instruction on how to ask questions, how to record responses, and how to handle any problems that may arise during data collection with CAPI. This includes ensuring the data collectors understand clearly the concept of informed consent, both oral and written while administering the survey. Field tests should be part of the enumerators training.
6. Pilot the survey. Run the survey test with a small group of respondents to identify any problems with the survey instrument, survey administration, or data collection procedures using CAPI. Apply any necessary corrections and repeat the pilot test until a sufficient degree of confidence and data quality is built.
7. Conduct the survey. Deploy the survey via face-to-face interviews with the selected sample using your preferred tool. Data collection staff should follow the survey protocols closely to ensure consistency and avoid bias. This includes taking a soft approach when asking sensitive questions to the respondent regarding their children, livelihood, well-being and lifestyle.
8. Quality control. Implement quality control procedures, for example, using your software administrator's dashboard to ensure the data collected is accurate and reliable - almost in real-time. This may also include spot-checking completed surveys, re-interviewing respondents (also known as back-checking) and checking data for missing or inconsistent responses in the dataset (also called high-frequency checks).
9. Data transmission. Each day of data collection, make sure all data are properly synchronized and available on the cloud server. It is recommended that the data for every session is synchronized after each session is performed.

#### TIP

Ensure that a research supervisor is available to monitor and check the data collection process while also recording with a watch how long a data collection session lasts. This is important because it provides an average of how many sessions an enumerator can do in a day and facilitates planning the data collection exercise. The research supervisor should also make sure that all the sessions recorded using the tablet are synchronized and uploaded to the cloud for every personnel collecting the data.

## 2.2 Analyze survey data

The first step in data analysis is careful planning. Even before collecting the data, the researcher must have the overall purpose of the exercise in mind. Developing a data analysis plan can be one way to organize and analyze the survey data. It should help you achieve three goals related to the purpose you set before starting the survey.

- Answer the main research questions
- Use more specific survey questions to understand the responses
- Sample survey participants to study different demographic groups

That said, when it comes to collecting data using wearable technologies, face-to-face survey data can be useful for other reasons, too. When used in combination with more sophisticated data collection, such as that based on wearable devices, reported information becomes very rich because it can provide indications on those social dimensions that could explain, at least in part, the patterns detected with the devices.

In other words, although reported survey data on child development should be used with caution, combining this information with technology-based data provides a useful tool. Many of the variables that could be generated using reported information on children's cognitive and social-emotional skills, parental beliefs and attitudes, and children's health and education could have explanatory power for the patterns we observe with the technological tools we use.

For example, a recent study using electroencephalogram (EEG) data found impacts of a poverty reduction intervention on children's brain activity and cognitive abilities (Troller-Renfree et al. 2022). The different dimensions of the poverty reduction intervention, as measured by the reported information collected through the surveys, explained the changes in cognitive development measured by the EEG data. This is the spirit that should be kept in mind when planning survey data collection in the context of a wearable device-based study.

Preparing survey data for such an analysis means paying attention to how participants are sampled in the study and ensuring that each data point generated is associated with a unique individual, such as a child identification (ID) number. Also, although binary variables are sometimes more convenient to manipulate, it can be useful for each dimension described through the reported information, such as cognitive abilities, attitudes, or parental activities, to generate normalized values or indices that one could then use to study correlations with trends in wearables-based data. A very simple example would be to study the correlation between language development (as measured with audio recordings), with gender and children's age (a binary and a scalar variable, respectively).

A word of caution should be spent on the analysis of the directly observed measures of child development. There are several survey instruments that could be used to measure the various developmental stages a child goes through in the early years of life. Two dimensions need to be balanced. On the one hand, nationally validated instruments might be better suited to the local and cultural context, especially when the study takes place in a developing country. On the other hand, internationally validated tools, less adapted to the field but comparable on an international scale, could be considered.

Regarding the analysis of these data, it must be noted that validated instruments always provide instructions on how to statistically generate the indicators of child development that they intend to measure. For example, the IDELA tool (<https://idela-network.org/>) comes with a toolkit for researchers to help navigate through the data collected to generate indicators related to four domains: motor development, emergent literacy, emergent numeracy, and social and emotional development.

# 3

---

## Phone-based survey



**Authors:**

Onicio Leal Neto, Daniela Paolotti,  
Nicolò Tomaselli, and Guilherme Lichand

## 3. Phone-based survey

Phone surveys are a type of data collection method in which a researcher conducts a survey by asking respondents questions over the telephone. Phone surveys can be conducted in a variety of ways, such as by manually dialing numbers and conducting interviews or by using an automated system in which respondents answer pre-recorded questions by pressing numbers on their telephone keypad, also known as Interactive Voice Response or IVR. In the Child Development Study in Malawi, phone surveys were termed “participatory surveillance” because participants have been asked to provide information over the phone, primarily about the health status of their children.

In public health, the practice of engaging the public and placing them at the center of elucidating new diseases and improving prevention is not entirely new. Participatory surveillance has shown positive results in different settings and events and has been presented as a complement to traditional health information systems. The use of collaborative information channels allows for more rapid collection of data on symptomatic individuals in the territory and offers the potential to anticipate epidemics or provide a quicker overview and guide authorities to the disease hotspots in question. Although this strategy has shown some limitations in developing countries, where access to the internet and technological devices is a constraint, patient participation in prevention and treatment is a global trend, and health care providers are incorporating techniques into their daily practice to share responsibility and decision making for the treatment plan.

History shows that community engagement was at the epicenter of one of the most iconic scientific discoveries in public health, when in the 1850s, John Snow elucidated the role of water in the transmission of cholera by incorporating reports from those affected by the epidemic into his investigation. Similar participation and engagement is now the core principle of participatory surveillance strategies that have been applied around the world to screen for pandemics such as Zika, influenza, and Covid-19.

The strategy can be built on accessible systems and applied in regions considered to have few resources, such as Malawi. By taking advantage of the Interactive Voice Response (IVR) design, a less expensive, flexible, scalable, and reliable system can be implemented that collects data on a voluntary basis and provides information that is not available using traditional monitoring methods. This allows communities to play an active role in anticipating outbreaks, even in places with limited phone or internet coverage.

### 3.1 Collect phone-based survey

Researchers may conduct multiple waves of telephone surveys, that is, repeating the survey weekly or monthly for a defined period of time. Data collection can be done by calling respondents and asking them questions. Another possible technique is the use of an automated system in which respondents

answer prerecorded questions by typing numbers on a telephone keypad (IVR). Here we focus mainly on participatory surveillance implemented with an automated system.

Data collection can be divided in two phases:

1. **Registration:** The researcher needs to collect the phone numbers of the participants who will be involved in the data collection. This data collection process will also include the collection of demographic characteristics, informed consent, and other crucial metadata of the participants.
2. **Data Collection:** The collected phone numbers are then uploaded to the software database, which has a created stream and is triggered to begin making phone calls with interactive voice responses (IVR) that ask the respondent questions from the questionnaire. These responses are then aggregated and exported to the dashboard where decision-makers can monitor data trends.

To begin data collection, the following items and services must be available:

1. **Phone survey software.** You need software that allows you to conduct telephone surveys. At the time this user-guide was written, for example, the companies Textit, Viamo or Twilio are among the cloud-based communication firms that provide programmable communication tools to make and receive phone calls, send and receive text messages, and perform other communication functions using its web service APIs.
2. **Phone numbers of survey participants.** They can be obtained through preliminary data collection in the survey area and households. The administrative system in place, such as the health district, can also provide this information at a lower cost.
3. **Survey.** You will need to develop a 10-15 questions instrument with relevant questions to collect the data. Make sure that the questions are clear, concise, and unbiased.
4. **Computer-based survey applications.** This interface can be used by interviewers to collect telephone numbers, metadata, and demographic characteristics
5. **A dashboard will be needed to visualize the flow of data from the survey response.** At the time this user-guide was written, the platform called Exploratory is a good example of such a software.

## BOX

Sensitisation meetings can be held in each of the communities of the study, prior to start phone-based data collection. The content of the meetings can be as follows:

- 1- Demonstration of a live call;
- 2- Presentation explaining in simple terms how the data will be collected and the relevance of the study to the child's development and overall health;
- 3- Clarification and reinforcement that the calls would be made weekly from the same number, asking the same questions.
- 4- Presentation explaining the importance of prevention and early detection of epidemics and empowering the participants as key collaborators for the success of such initiatives;

5- Reinforcing the possibility to call back and continue to participate in the current week, in case the participant missed a call. The procedure was available at no cost to the participant and always at the same local number.

6- Request for a preferred time slot to receive calls.

7- Open time for participant questions and requests for further clarification.

This approach clarifies any technical issues regarding the use of the cell phone in this specific context, and reinforces that the repeated calls are not a mistake, but a normal and essential procedure for the success of the study. Last, it raises participants' awareness of the study and the importance of their participation.

Data collection can be divided in two phases:

- Registration: The researcher needs to collect the phone numbers of the participants who will be involved in the data collection. This data collection process will also include the collection of demographic characteristics, informed consent, and other crucial metadata of the participants.
- Data Collection: The collected phone numbers are then uploaded to the software database, which has a created stream and is triggered to begin making phone calls with interactive voice responses (IVR) that ask the respondent questions from the questionnaire. These responses are then aggregated and exported to the dashboard where decision-makers can monitor data trends.

## 3.2 Analyze phone-based surveys

Interactive voice response (IVR) systems can facilitate participatory surveillance by enabling efficient and cost-effective data collection. In many studies, frequent automated phone calls are made to a representative sample of the population over a period of time, during which participants answer a consistent set of questions related to a specific topic and associated risk factors. As a rule of thumb, a response rate of less than 50 percent should be expected, with a phone-survey completion rate of less than 30 percent. The awareness campaign and SMS reminders could greatly increase these figures.

The data can be analyzed in different ways. For simplicity, we discuss here below how in low-income countries and in settings where technological resources are scarce, participatory surveillance can be particularly useful for the prevention of disease outbreaks.

1. Analysis of aggregate data. What is special about this type of data and data analysis is that it is not necessarily based on individual identification. In other words, the dimension of analysis is not the individual or the household, but the aggregate value at the community level. Data analysis focuses on trends and not individual correlations. The relative anonymity offered by an IVR system can also reduce social desirability bias by encouraging participants to provide more accurate information on sensitive topics such as symptoms or risk behaviors.
2. Longitudinal information. By following the same group of participants over a sufficiently long period of time, the IVR-based participatory surveillance system captures temporal trends and identifies emerging risk factors, enabling targeted and context-specific interventions to prevent and control outbreaks. The longitudinal dimension should be taken into account when preparing data for

analysis. Various econometric techniques can be employed to extract knowledge from longitudinal panel data or repeated cross-sectional data.

3. Case detection. The IVR system enables rapid identification of suspected cases within communities, providing real-time data to respond quickly and contain potential outbreaks. To this end, the analyst must focus on the spikes that may appear in the data compared to the average observations in the dataset. For this reason, it is important to automate the updating of the threshold against which the data are compared in each time period. When a data point exceeds the threshold by a certain value, that data point is the trigger for an alarm.

Limitations of analyzing this type of information may arise when population shifts and changes in household configuration occur. This can affect the consistency of individuals reported within the same household phone number. In some cases, continuous changes in phone numbers may result in the loss of contact of an entire household. More generally, the analysis should examine problems with adherence rates over the course of weeks, as receiving the same phone call each week with the same questions may cause a person to drop out of the survey, e.g., selective attrition.

### Useful references:

Wójcik, O. P., Brownstein, J. S., Chunara, R., & Johansson, M. A. (2014). Public health for the people: participatory infectious disease surveillance in the digital age. *Emerging Themes in Epidemiology*, 11(1), 7. <https://doi.org/10.1186/1742-7622-11-7>

Leal Neto, O., Cruz, O., Albuquerque, J., Nacarato de Sousa, M., Smolinski, M., Pessoa Cesse, E. ., Libel, M., & Vieira de Souza, W. (2020). Participatory surveillance based on crowdsourcing during the Rio 2016 Olympic games using the Guardians of health platform: Descriptive study. *JMIR Public Health and Surveillance*, 6(2), e16119. <https://doi.org/10.2196/16119>

Ruikar, V. (2016). Interactive Voice/Web Response System in clinical research. *Perspectives in Clinical Research*, 7(1), 15–20. <https://doi.org/10.4103/2229-3485.173781>

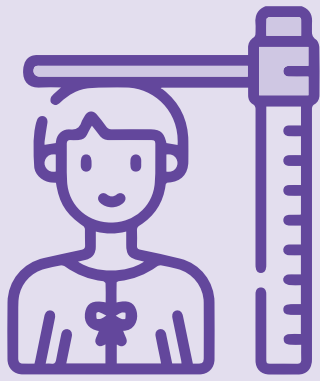
Marcano Belisario, J. S., Jamsek, J., Huckvale, K., O'Donoghue, J., Morrison, C. P., & Car, J. (2015). Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. *Cochrane Database of Systematic Reviews*, 2015(7), MR000042. <https://doi.org/10.1002/14651858.MR000042.pub2>



# 4

---

## Anthropometric measurements



**Authors:**

John Phuka, Beverly Laher, Lonjezo Dzekedzeke,  
Nicolò Tomaselli, and Guilherme Lichand

## 4. Anthropometric measurements

Measuring a child's physical development is an essential component for studying child development. Anthropometric measurements are a collection of physical standards used to evaluate a person's size, shape, and body composition. These measurements are applied in a variety of contexts, such as clinical settings, public health programs and research on nutrition and medicine. Accurate anthropometric data is also important for nutritional assessment, and ergonomic design. Collecting anthropometric measurements requires standardized protocols and appropriate measurement tools, such as scales, stadiometers, and tape measures. One can measure metrics such as height, weight, waist circumference, mid arm circumference, head circumference and skinfold measurements. This section will focus on measurements commonly used in measuring child growth and development.

### 4.1 Collect anthropometric measurements

To collect the anthropometric data, the following steps should be undertaken. Prepare the child before taking any measurements, explain to the child what will happen and make sure they are comfortable and relaxed. If possible, have the child wear light clothing and remove any bulky items. All anthropometric measurements should be performed by two trained persons, an examiner and a recorder.

- 1. Weight measurement.** Have the child stand on the scale with their feet shoulder-width apart and their arms at their sides. Record their weight to the nearest 0.1 kg. When possible, use Secca brand scales. Children under 2 years of age or who cannot stand still are best weighed with the mother holding them, e.g., tared weight. In this situation, weigh the mother first, then turn the scale to the tared mode and weigh the mother and child together; the scale will then display the weight of the child. Use the tared weight if the youngster is unable or unwilling to stand on the scale. Children who are two years old and older can be weighed by themselves as long as they remain still. It is preferable to use the tared weighing method if the child is restless. It is advised that children be weighed with little or no clothing. Some parents or caregivers might not permit the child to be weighed without clothes due to cultural considerations or the environment. Children may be blanketed to meet this preference. First ask the adult to stand on the scale with the blanket and tare the scale so that the weight of the blanket used to cover the child while weighing will not be included when measuring the child's weight. The adult should then wrap the child in the blanket and stand on the scale while holding the child for measurement.
- 2. Circumference measurements.** Use a non-stretchable standard Mid-Upper Arm Circumference tape, also Known as a MUAC tape, to measure the child's head circumference, mid-upper arm circumference, waist circumference, and hip circumference. For head circumference, place the tape measure around the largest part of the head, just above the eyebrows and ears. For mid-upper arm circumference, measure the midpoint between the shoulder and elbow on the upper arm. For

waist circumference, measure at the natural waistline, which is the narrowest point between the ribcage and the hipbone. For hip circumference, measure at the widest point of the hip. Record each measurement to the nearest 0.1 cm.

Head Circumference. Depending on their age and amount of activity, the child either stands, sits on the footstool, or sits in the parent's lap. The insertion tape is wrapped around the head just above the ears on either side, across the frontal bones just above the eyebrows, and over the occipital prominence towards the rear of the head. The examiner tightly wraps the insertion tape around the subject's head. To find the maximum circumference of the head, the insertion tape is moved up and down over the back of the skull. In order to compress the hair and underlying soft tissues, the insertion tape must be pulled tightly while being perpendicular to the long axis of the face.

Waist circumference. The lateral border of the ilium is a bone landmark that must be found and marked in order to determine the level at which the waist or abdominal circumference is measured. Place the SP on their feet as they hold the examination robe above their waists. While standing behind and to the right of the SP, loosen the SP's garments and palpate the hip region to determine where the right hip should be. The midaxillary line of the body is indicated by drawing a horizontal line crossing it slightly above the topmost lateral border of the right ilium. Place the measuring tape around the trunk in a horizontal plane at the level indicated on the right side of the trunk while standing on the right side of the child.

Mid-arm circumference. The individual should be standing straight, with relaxed shoulders, and the right arm should be dangling lightly. It's crucial to make sure the arm is not flexed or tensed, as this could result in a greater and erroneous reading. Place the measuring tape around the upper arm at the crossing point (+), perpendicular to the long axis of the upper arm, while facing the SP's right side. On the skin's surface, hold the measuring tape gently. Measure the lateral aspect of the arm by pulling the two ends of the overlapping tape together while maintaining the zero end below the measurement value.

3. Height measurement. Have the child stand straight against the stadiometer with their heels, buttocks, and back touching the wall. Make sure their head is in the Frankfort plane (ear and eye level are equal) and their arms are at their sides. Record their height to the nearest 0.1 cm. When possible, use Secca brand stadiometers with a vertical board and adjustable head piece. Children aged two and above, who can stand unassisted, can be measured by placing the equipment on a flat even surface. The backboard should ideally rest on a wall. This measurement is taken while standing, with the individual's shoes off.

The mother or guardian should remove any hair ornaments, or braids from the top of the head of the child before measurement is taken. Remember to ask the participant to stand up straight against the backboard with the body weight evenly distributed and both feet flat on the platform. Instruct them to stand with the heels together and toes apart. Check that the back of the head, shoulder blades, buttocks, and heels make contact with the backboard. The examiner must ensure that the arms and shoulders are relaxed and held on the side of the body, the head should lie in a Frankfort plane (examiner may need to physically assist in positioning the participant's head). The head is in the Frankfort plane when the horizontal line from the ear canal to the lower border of the orbit of the eye is parallel to the floor and perpendicular to the vertical backboard.

Recumbent length. Be aware that this technique is suitable for children aged two years and below, but can be used on children up to 4 years. Once the participant is seated properly, lower the

stadiometer headpiece until it is firmly resting on top of the participant's head and has applied enough pressure to compress the hair. Give the participant instructions to stand as tall as they can, take a deep breath, and maintain their posture. Note how the headgear tilts slightly during breathing. The examiner will then call the reading to the recorder. Whenever possible, use the Secca Infantometer. The equipment should be placed on a flat, even surface. The measuring table has a fixed head, a horizontal back part, and a movable foot part. When measuring length in the supine position, the parent or caregiver should be positioned between the examiner and the recorder. The parent should encourage and comfort the child by making eye contact, talking and, if necessary, holding a restless child's head. The recorder supports the child's head and places it on Frankfort's floor. Gentle traction is applied to bring the top of the head into contact with the fixed earpiece. The child's head should be held firmly in this position by gently cupping the palms of the hands over the ears and keeping the head in proper alignment. Simultaneously, the examiner aligns the child's legs by placing one hand gently but firmly on the knees. The toes point directly upward with the soles of the feet perpendicular to the horizontal back of the measuring device. The knees are pressed gently to keep the legs straight. When the child's heels are firmly in place, the examiner slides the movable piece of the foot to rest it. The examiner will position the child as indicated above and use the tape measure installed on the infantometer board to read the length if there is a power failure or if the infantometer is not working properly. The recorder will write down the measurement once the examiner has told him or her.

*Note: Placing infants and small children in a recumbent position frequently generates a sense of insecurity and consequently invokes a crying response. Mothers and Guardians play an important role in keeping the child calm and getting an accurate reading.*

Anthropometrists' preparation is crucial. Prior to approaching the child who will be measured, the recorder and the examiner must both have clean hands and short fingernails. Any items on their hands or wrists, such as bulky watches or bracelets, must be taken off in order to avoid impeding the measurement or potentially harming the child. Placement of the measuring board and scale should be considered carefully. Make sure there is a stable, flat surface for the measuring board and digital scale, as well as enough light to see the measurements precisely. Consider bringing a wooden board to level the scale if the ground is not level.

The anthropometrist should explain each step to the child and the mother or guardian before starting it. Never leave a child alone with a piece of equipment; always remain in close proximity to the child except for the short period necessary to take the child's weight. Remember that the assistant can and should talk to and comfort the child during the measurement procedure, but should not participate in it.

### **BOX:**

Individuals with disabilities. It is recommended to measure individuals with disabilities. Accurate and safe measurements, however, can be difficult to obtain in people who are unable to stand, straighten their arms, legs, or back, or maintain their balance. In certain situations, it could be necessary to modify the measurement procedures or provide the child being measured more support.

## 4.2 Analyze anthropometric measurements

Generally, anthropometric measurements are used to determine a child's nutritional and general health status. Measuring the weight and height of infants and children is an international health practice that provides a readily available and objective method for determining a child's history and health status.

The WHO Anthropometric Survey Analyzer promotes best practices in the analysis of anthropometric indicators. It analyzes five indicators: length/height for age, weight for age, weight for length, weight for height, and body mass index for age. In large samples, anthropometric characteristics tend to be continuous, and many tests are developed under the assumption that the data follow a normal distribution. An easy way to see if the distribution is skewed is to compare the mean and median values. For normal distributions, the mean and median should be identical. As the distribution becomes more skewed, the difference between the mean and median increases. There are several statistical tests for testing normality and the researcher may get different results depending on the test used. For example, the Kolmogorov-Smirnov test examines the cumulative distribution, which confounds skewness and kurtosis, while the Cox test determines the extent of skewness and kurtosis separately.

A simple tip for researchers wishing to analyze anthropometric data is to refer to the WHO procedures and software, which at the time this user-guide was written could be found at the following link: <https://www.who.int/tools/child-growth-standards/software>.

The WHO methodology is updated to reflect the sampling design of recent surveys. The improvements affect only the measures of precision around the estimates (i.e., standard errors and confidence intervals). The tool is based on R, with the Shiny package. Outputs include a set of z-scores, a file with prevalence estimates by the different stratification variables following the expanded database format, a report template on data quality assessments, and a summary report with a fill-in-the-blank template with the basic information required for the survey and ready-to-use graphs and tables describing the results of the survey analysis. Both online and offline versions of the tool can be accessed from the links above. Practitioners are encouraged to read the quick guide carefully before using the tool.

### **BOX:** *what is a z-score?*

In simple terms, a z-score (also called a standard score) gives an idea of how far a data point is from the mean. More technically, it is a measure of the number of standard deviations below or above the population mean of a raw score.

A z-score can be placed on a normal distribution curve. The z-scores range from -3 standard deviations (which would correspond to the extreme left of the normal distribution curve) to +3 standard deviations (which would correspond to the extreme right of the normal distribution curve). To use a z-score, you need to know the mean  $\mu$  as well as the standard deviation of the population  $\sigma$ .

# 5

---

## Language interaction

**Authors:**

Adriana Weisleder, Casey Lew-Williams, Paola Garcia, Asana Okocha, Jessica Kosie, Alexandra Carstensen, Nicolò Tomaselli, and Guilherme Lichand

## 5. Language interactions

One relatively new source of information on child development comes from the development of speech capacity and child's verbal interactions with peers and family members. In practice, language and verbal interactions refer to the exchange of information and ideas between individuals or groups through spoken or written language. Effective language and verbal interactions requires children to understand and use language appropriately in order to convey messages and understand the messages of others. One way to gather information about language interactions is to use portable devices that can record voices with microphones. Although this method seems quite simple, it is the analysis of this information that is the main innovation. Indeed, modern technology allows the use of software to process the recordings and generate indexes that can be used for analysis.

More generally, the investigator wants to collect a variety of data to characterize the child's early language environment and measure language development. These streams of data can be separated into behavioral data, assessments, and self-report.

1. Behavioral data includes daylong audio recordings in which we capture all language used by the target child and the language spoken by adults and children in his or her vicinity. We could also record a structured play session in which a caregiver interacts directly with the target child for around 10 minutes. From these recordings, we can calculate key variables including adult word count, conversational turns, and child vocalizations.
2. Standardized assessments using validated tools could be used to measure a child's language development in a given country, e.g. Malawi. For children 36 months and below, one could use the MacArthur Bates Communicative Development Inventory (MCDI; Fenson, 2007), and for those children above 36 months, one could use the language section of the Malawi Development Assessment Tool (MDAT) or any other validated tool adapted to the context.
3. Finally, self-reported demographic data provides further contextual information that may aid in describing children's early language environments. For example, household composition, family structure, household wealth, and caregiving arrangements may meaningfully predict caregivers' language use and ultimately, the child's development.

### 5.1 Collect language interactions data

Audio recording is a media for capturing the rich dynamics of human conversation in language and speech interactions research. However, it takes careful preparation and execution to gather high-quality audio data. Here below, we outline the procedures for obtaining audio data from recorders for use in language and speech interactions research, including choosing the right recording tools, setting up the recording environment, and managing recorded data.

For short recordings of natural interactions, any recorder can be used, including a phone or tablet. For multi-hour or daylong recordings of natural interactions, a portable audio recorder should be inserted into the clothing of the child, to capture speech occurring near the child.

For assessments of child vocabulary, one could administer parent-report vocabulary checklists such as the MacArthur-Bates Communicative Development Inventory, or CDI, adapted to the language of interest. Finally for demographic and household data, one could administer questionnaires that ask parents about verbal-interactions-specific aspects of their lives and backgrounds.

In order to start collecting the recordings, you need to have at least the following equipment.

Recorders

Vests (different sizes available for different ages)

Cling wrap

Tablet with computer assisted personal interviewing software

Batteries

Rubber or plastic to lock the recorder with

Toys (optional)

Timer/watch

Before beginning the data collection process, it is recommended that each enumerator have a hard copy of a tracking sheet. This document will be used to identify households selected to participate in the data collection exercise. The tracking sheet should contain all the essential details needed to identify the household, such as the household ID, district, village, location, and other information of those selected to participate in the data collection. In addition, the enumerator could ask the guardian to provide a passport or health certificate containing verifiable information about the participant's identity.

To start, make sure the tablet and the digital recorders are charged and that memory space is available. Also, make sure the time and date settings are accurate before starting a session. Select a clean and right fit cloth vest for the child to be used during the session.

Next, have a verbal consent script at hand. Select a quiet place where you will conduct the session. In the following example, we describe a procedure divided in two parts. The first part consists of a recording of more or less 10 minutes, which involves the interactions between the guardian and the child. Here is how it works.

For each recorder:

1. Turn on the recorders
2. Clap your hands once (to know when the recording has started and to synchronize the second recorder).
3. Say aloud the date and time, household ID, child ID, and your name.
4. Place a piece of plastic or rubber over the "HOLD" button, so the child cannot turn off the recorder.
5. Take a piece of cling film and wrap it around the recorder several times, making sure not to cover



the microphone (top of the recorder). This is important because it prevents children from changing the settings.

6. Slide the recorder into the vest pocket with the microphone at the top of the pocket, then snap the pocket closed. Put the vest on the child, making sure it fits snugly.

After this preparatory phase, ask the guardian to put the vest on the child. Accompany the child and parent to a place where you think it is best to check in. Ideally, this is a quiet place away from other children.

Tell the parent, "We are going to record you and [child's name] for 10 minutes while you play, doing the activities you normally do. For example, you can talk to each other, pretend to buy something at the market, roll a ball back and forth, pretend to cook together, spin a top, or encourage the child to walk or do another activity." If available, a toy can encourage the child to engage more with the parent.

Establish a perimeter where the exercise will take place. Say, "We want to make sure we can hear you playing with [the child], so speak as you normally would. I'll be back in 10 minutes." Set a timer for 10 minutes.

After the recording, take notes on your CAPI application to answer the questions below:

---

Where did the recording take place?

---

How were the parents and child sitting? (facing each other; side by side; child on lap...)

---

Who else was present during the recording?  
(include all children and adults, with approximate child ages)

---

The second part of the recording is less controlled. To start with, tell the parents that their voices will be recorded for about 12 hours using a digital recorder. Explain that the device is to be left on and recorded for many hours and that they are to turn it off when the child falls asleep. Make sure that both the parent and child are comfortable with the recording. Remind them to let everyone around the child know they are being recorded. Let them know what time you will pick up the check-in the next day.

When picking up the recorder, please ask these questions and record the answers:

---

Did anything unusual happen during the day of the recording? (For example: Was one of the children sick? Did you have different visitors than normal? Did you do different activities than normal?)

---

Who was present during the 12-hour recording? (Give your best guess; include all children, adolescents, and adults)

---

Once you have completed both registrations, please add any additional notes about your experience with the family, in particular for details you don't want to forget.

When collecting the recorder, place the recorders belonging to one household in an envelope and write down the metadata about the household on a piece of paper, then indicate which recorder belongs to a particular child. This is important because it is a safeguard in case the enumerator did not correctly specify which recorder was given to a particular child or if the parents switched recorders for the children. This also helps to create folder names when downloading data from a large number of recorders belonging to different households.

Then, to download the speech and language development data, connect the recorder to a USB port on a computer. Once connected, the computer will detect the files in the recorder and prompt you to open them. The recorder should contain at least two recording files, one 10 minutes long and one hours-long. Create a folder on your computer where you will store the recordings and give the folder an appropriate name. Within this folder, create a subfolder in which you will store the recordings for each participant. For example, if the participant's household ID is 553133, name the subfolder "553133."

If you are collecting records for two participants in the same household, create a separate subfolder for each child within the main folder. To label these subfolders, use the following example: For the first child, name the subfolder "553133\_1". For the second child, name the subfolder "553133\_2". Cut the recording files from the recorder and paste them into the sub-folder you created in step 3. When you paste the files, be sure to rename them using the following format: For the 10-minute recording, name the file "553133\_1\_rec1". For the 12 hour recording, name the file "553133\_1\_rec2". Once the download is complete, you will need to upload the data to the cloud server of your choice. Make sure the data is encrypted to maintain participant privacy. Following these steps will ensure that the data is organized and easily accessible for later analysis.



## 5.2 Analyze language interactions data

Data processing involves the application of signal processing and artificial intelligence techniques. Initially, the researcher must detect and eliminate empty files, since these are spontaneous speech recordings obtained from a USB recorder. However, due to the nature of these recordings, additional cleaning and processing is required. In the next step, periods of silence must be removed from the recordings, so that only speech segments remain. The resulting speech segments are then classified into different categories, including

- the speech of the target child,
- the speech of other participants in the conversation, and
- the speech of the far field, which is usually excluded.

Once each participant's speech has been separated, we can examine other characteristics, such as turn of speech, child vocalization, and number of words spoken.

The following parameters are usually measured.

1. **Word choice and vocabulary.** The words and phrases used during communication can have a significant impact on how the message is received. Measures of word choice and vocabulary can include measures of complexity, variety, and appropriateness of language.
2. **Grammar and syntax.** Correct grammar and syntax are essential for effective communication. Indicators related to grammar and syntax may include measures of sentence structure, use of tenses and voice, and presence of errors
3. **Clarity and coherence.** The clarity and coherence of a message are important factors in determining how well it is understood by the audience. Measures of clarity and coherence can include measures of organization, logical flow, and use of transitions.
4. **Engagement and Participation.** Effective communication involves the engagement and active participation of all stakeholders. Indicators of engagement and participation may include measures of listening skills, turn taking, and use of questions and comments.

Obviously, depending on the language spoken at home, the software chosen will be able to perform some tasks and not others. However, we remain convinced that the study of language development is essential to ensure that the various dimensions of child development are taken into account in the design of support policies.

### Useful references:

<https://www.cambridge.org/core/journals/journal-of-child-language/article/combining-observational-and-experimental-approaches-to-the-development-of-language-and-communication-in-rural-samples-opportunities-and-challenges/939868E21287B5C84EAF03D374463FEC>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5510534/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7337141/#R89>

[https://link.springer.com/epdf/10.3758/s13428-020-01365-9?author\\_access\\_token=ZxR3BfkXqc9IXhomWjYCzJAH0g46feNdc402WrhzyrTEdBmrm49GxadLGuBXz0GUi32pgoTnQqySZGmi6uefKRSvcGr8RqKciERhSwmt0tYubA7xuBGp7RSJ6SfQ2U74XfIXKt9jPAhdO1x8mVEg==](https://link.springer.com/epdf/10.3758/s13428-020-01365-9?author_access_token=ZxR3BfkXqc9IXhomWjYCzJAH0g46feNdc402WrhzyrTEdBmrm49GxadLGuBXz0GUi32pgoTnQqySZGmi6uefKRSvcGr8RqKciERhSwmt0tYubA7xuBGp7RSJ6SfQ2U74XfIXKt9jPAhdO1x8mVEg==)

[https://psycnet.apa.org/record/1995-98021-000?mod=article\\_inline](https://psycnet.apa.org/record/1995-98021-000?mod=article_inline)

[http://www.uh.edu/class/psychology/dcbn/research/cognitive-development/\\_docs/mcdigestures.pdf](http://www.uh.edu/class/psychology/dcbn/research/cognitive-development/_docs/mcdigestures.pdf)



---

Social  
interactions



**Authors:**

Ciro Cattuto, Lorenzo Dall'Amico,  
Nicolò Tomaselli, and Guilherme Lichand

## 6. Social interactions

An even more innovative measure is close-range proximity interactions between individuals. Tracking human-to-human proximity can provide critical information about social contact in communities and related social behaviors linked to health and development outcomes. One way to measure social interactions is through wearable proximity sensors, which can measure the distance or detect the proximity and co-location of individuals wearing them.

Wearable proximity sensors can use several technologies to detect close-range contacts. Popular approaches employ a variety of low-power radio technologies, including Radio-Frequency IDentification (RFID), Bluetooth or Bluetooth Low Energy (BLE), Ultra-Wide Band (UWB), custom radio protocols, etc. The sought temporal and spatial resolutions largely determine the most suitable technology for a given application, and in turn the power requirements of the sensors, hence constraining the battery duration. Data extraction can similarly be carried out in different ways that span LoRa (Long Range radio), Bluetooth radio transmission to data sinks, or manual data extraction via dongles that make physical contact with the sensors.

Several techniques have been successfully demonstrated for small-scale pilots and data collections. Few systems have demonstrated to scale well to thousands of sensors – the scale sought here. Scalability in practical applications entail a workable, and often challenging, combination of low unit cost, long battery duration, reliable operation, adequate non-volatile local storage, and usable toolchains designed for non-expert operators. For the Child Development Study in Malawi, researchers leveraged the proximity sensing system and toolchain developed by the SocioPatterns collaboration ([www.sociopatterns.org](http://www.sociopatterns.org)) led by ISI Foundation in Turin, Italy. The system has been developed over the course of a decade and has been deployed in a variety of environments that spanned schools, hospitals, social gatherings, museum, households, and challenging large-scale deployments in low-resource settings ([www.sociopatterns.org/publications](http://www.sociopatterns.org/publications)).

It is important to note that proximity sensors do not collect geographical location information, such as GPS receivers. Hence proximity sensors provide, if properly deployed, higher guarantees of data protection, although care should be taken with empirical social network data since it is established that social graphs are basically impossible to anonymize. It is also important to remark that these data provide objective and high-resolution information on a rather specific type of social behavior (proximity in physical space), but they cannot provide contextual information that is often crucially needed to make sense of proximity interactions (e.g., whether the interacting individuals were speaking, body posture and movement, etc.). Hence, these data are best suited to study phenomena for which close-range proximity is a known determinant (e.g., spread of an airborne pathogen), time spent together by individuals in a household, etc.

## 6.1 Collect social interactions data

The dynamics of social behavior exposed in close-range mobility and proximity can be effectively investigated by gathering proximity data using wearable sensors. Practitioners and researchers can study people's contacts, closeness and co-location using proximity sensors. To ensure the accuracy and reliability of the collected data acquired, meticulous planning and execution are necessary when gathering social interactions data with proximity sensors.

In the Child Development Study in Malawi, the SocioPatterns proximity sensors were deployed in rural villages. The sensors use low-power radio packed in the ISM (Industrial, Scientific Medical) radio band, and a much simpler protocol than Bluetooth. The exchange of low-power radio packets and the thresholded attenuation (RSSI) of the signal is used as a proxy for the close-range proximity of individuals wearing the sensors. The system achieves a spatial resolution of 1.5-2m, a temporal resolution of about 20 seconds, battery duration in excess of 3 weeks, and a low unit cost that easily affords scaling up to several thousand sensors with the financial resources of a research study. The SocioPatterns sensors (shown below) were protected by custom-designed 3D-printed plastic enclosures.

Data extraction in the Child Development Study occurred offline, after collecting the sensors at the end of the study, and relied on a custom USB dongle that makes physical contact with the sensors and extracts the data to a Web app running in the Chrome browser. The extracted data files, one per sensor, are pooled by the operators in the field and transferred to a centralized secure location for data cleaning and post-processing. The data collected by the sensors contain proximity information, but the sensors do not apply any specific contact definitions to the data, leaving the specific contact definition – in terms of attenuation and temporal smoothing – to the researchers analyzing the data, so that the best contact definition for a given research objective can be used. The collected data also contains timestamps, low-frequency information on the orientation of the sensors in space, whether they have moved recently or not, the battery level, and several diagnostic features.

The data post-processing pipeline is a stack of software components too complex to be described here. It consists of a pipelined series of post-processing stages that carry out the following operations:

- Analytics on data quality and sensor wearing compliance
- Temporal alignment and clock drift modeling
- Acceleration-based data filtering, to remove spurious contacts involving abandoned sensors
- Contact definition application and generation of temporal proximity network
- Linking between sensor hardware IDs and study IDs
- Data export to the scientific data analysis tasks

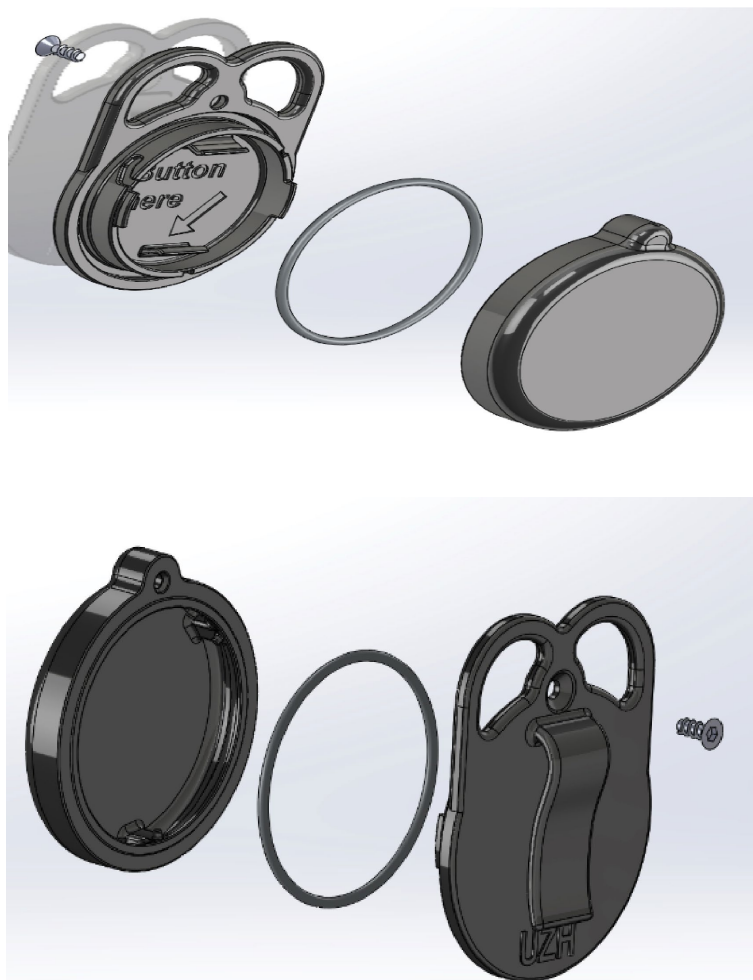
### Useful references:

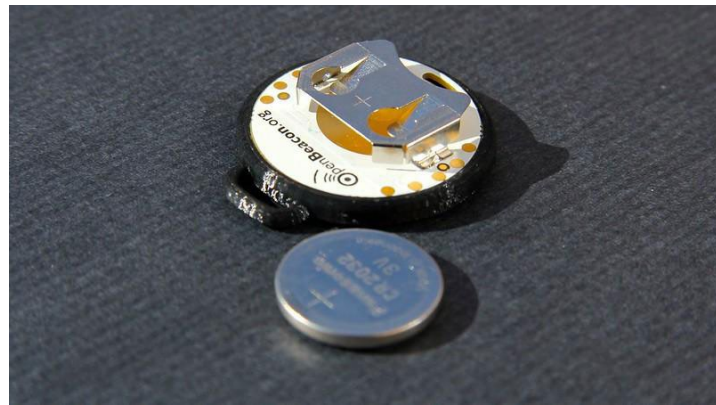
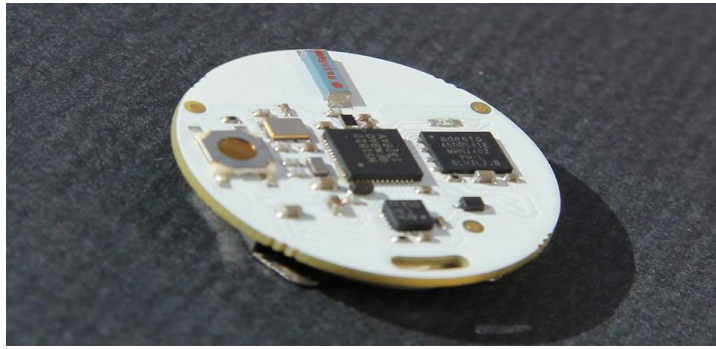
Barrat, A., Cattuto, C. (2015). Face-to-Face Interactions. In: Gonçalves, B., Perra, N. (eds) Social Phenomena. Computational Social Sciences. Springer, Cham. [https://doi.org/10.1007/978-3-319-14011-7\\_3](https://doi.org/10.1007/978-3-319-14011-7_3)

Kiti, M. C., Tizzoni, M., Kinyanjui, T. M., Koech, D. C., Munywoki, P. K., Meriac, M., Cappa, L., Panisson, A., Barrat, A., Cattuto, C., & Nokes, D. J. (2016). Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Science*, 5(1), 21. <https://doi.org/10.1140/epjds/s13688-016-0084-2>

Kiti, Moses Chapa, Melegaro, A., Cattuto, C., & Nokes, D. J. (2019). Study design and protocol for investigating social network patterns in rural and urban schools and households in a coastal setting in Kenya using wearable proximity sensors. *Wellcome Open Research*, 4, 84. <https://doi.org/10.12688/wellcomeopenres.15268.2>

Leal Neto, O., Haenni, S., Phuka, J., Ozella, L., Paolotti, D., Cattuto, C., Robles, D., & Lichand, G. (2021). Combining wearable devices and mobile surveys to study child and youth development in Malawi: Implementation study of a multimodal approach. *JMIR Public Health and Surveillance*, 7(3), e23154. <https://doi.org/10.2196/23154>









7

---

Socio-emotional  
development

**Authors:**

Michelle Bosquet Enlow,  
Nicolò Tomaselli, and Guilherme Lichand

## 7. Socio-emotional development

**DISCLAIMER:**

This user guide describes the collection of ECG data using a handheld device that we will refer to as the MAWI pad. These electrodes were manufactured by the company MAWI (<https://platform.mawi.band/>) and were used to collect heart rate and variability data. Because contracting with a vendor for this type of activity plays an important role in successful data collection, we provide in appendix an example of terms of reference that may inspire future conversations with similar tech companies.

In addition to the previous data categories, we now describe the collection of biomarker data. One marker, increasingly used in non-medical settings, is the measurement of autonomic reactivity, particularly cardiac activity, which can be used as a proxy for biological responsiveness to the environment and stress responsiveness and regulation, important factors in social-emotional development.

The ECG (electrocardiogram) measures the electrical activity of the heart, specifically the electrical impulses generated by the heart as it beats. From the ECG, heart rate and rhythm can be determined, as well as potential abnormalities in electrical activity that may indicate heart disease. The ECG is a non-invasive, painless test that is commonly used in cardiology to monitor various heart conditions.

In the field of psychosocial development in particular, these parameters may be used to contribute to predictions of clinical conditions that may affect child development, such as stress-related conditions, infectious diseases, or iron deficiency. These parameters may also be used as one indicator of the child's ability to physiologically regulate stress and other environmental inputs.

There are devices that collect data on cardiac activity in a manner that requires significant equipment to be placed on the child and heavy, time-intensive data processing of collected data. Another way to collect ECG information is through wearable devices that are, in contrast, easy to use. These devices use sensors that are stimulated by thumb pressure or, for children under 3 years of age, by wired clamps worn on the wrist. Through an automated process using software, the sensors collect and process the ECG data and generate indices that can be used for analysis.

From such devices, the following parameters can be measured, among others:

1. Heart rate. Measurement of the time between successive R-waves (the peak of the electrical signal in a heartbeat).

2. Heart rate variability. Includes time-domain measures, such as standard deviation of normal-to-normal intervals (SDNN), root mean square of successive differences (RMSSD), and percentage of successive normal-to-normal intervals that differ by more than 50 ms (pNN50); and frequency-domain measures, such as high frequency (HF) and low frequency (LF) power.
3. Conduction intervals. The duration of the various intervals in the cardiac conduction system, which are the pathways through which electrical signals travel in the heart. These intervals include the PR interval, which represents the time between atrial depolarization and ventricular depolarization, and the QT interval, which represents the time it takes for the heart muscle to contract and then recover.
4. Wave morphology. ECG machines can analyze the shape of the waves on the ECG tracing, which can provide information about the health and function of different parts of the heart, such as the atria and ventricles.

## 7.1 Collect electrocardiogram data

A structured protocol is necessary to collect accurate ECG data and ensure its consistency, reliability, and validity. The procedures and protocols described in this user-guide aim to provide accurate and complete instructions for ECG data collection, including electrode placement, recording techniques, and basic interpretation of results. This document describes ECG data collection supported by a handheld device that we reference here as the MAWI pad. These hand pads are manufactured by the company MAWI (<https://platform.mawi.band/>) and are used to collect data on heart rate and variability.

Children should be instructed to hold the pads in their hands with their thumb resting on a metal electrode for at least three minutes. For younger children with smaller hands, wire clamps have been used to connect the child's wrist to the electrodes. Wire clamps are also preferable for children who are unable to keep their thumb pressed on the metal electrode without interruption for the period of data collection to avoid interruptions or loss of data. The following procedure, while specific to MAWI pads, could easily be applied to other portable devices capable of monitoring ECG.

Before starting, the researcher should make sure to select a quiet and comfortable location to conduct the session and ensure to have these items available

Charged Tablet with Mawi App & Metadata collection software installed

MAWI pad

Kitchen VEG board

A chair

Power bank and Charger

Wrist bands

Electrodes

To collect the ECG data, the following steps should be undertaken:

1. Make sure the tablet and MAWI pad are fully charged.
2. Switch on the tablet. Note that the MAWI pad needs a software app to transmit data to the tablet via Bluetooth. Make sure the most updated version of the application is available.

3. Click on the app on the tablet and open the application.
4. In the case of the Child Development Study in Malawi, you would fill out Region, District and village name, in order to describe the setting in which the data collection is taking place.
5. Select the participant and fill in all the required details, which includes informed consent and household information. Thereafter, select the option where you are required to fill out the information about the guardian and the child. For the registration of the participants you will be required to add the participant's biographical data and you will assign the Participant ID for the child. Make sure the ID assigned to the child matches the ID captured on the metadata, to allow for data reconciliation.
6. The MAWI app will be connected to the tablet using Bluetooth. Connect the pad to the tablet by clicking on the serial number of the pad that has the same number as the one that has been detected by the tablet. Once this is done, go back to the participant profile and choose the name of the participant from whom you want to collect data.
7. If the child cannot hold the pad firmly, connect the wristband. Apply the red cuff to the right wrist joint and connect it with the red electronic cable and tighten it. Apply the yellow cuff on the left wrist with the yellow electronic cable and make sure the nut is tight enough. See pictures. The electronic cable should be connected to the portable pad.
8. Make sure the child's hands are resting on a kitchen veg board to reduce movements as much as possible.
9. Check that you can see the signal coming into the app. In the top part of the screen, you should see a red wave line with peaks (ECG).

As part of the study in Malawi, the researchers decided to stimulate the child to measure the responsiveness of the child's autonomic nervous system (ANS). In accordance with the usual procedures in this type of research, the application starts playing videos to stimulate the child. The sequence from the protocol is described below:

- i) Neutral video 1=120 seconds  
-10 second pause
- ii) Fear video=120 seconds  
-10 second pause
- iii) Neutral video 2=120 seconds  
-10 second pause

When the video playback is complete, the application will prompt to validate the result and synchronize the record. Immediately after data collection, the researcher must access the metadata form to complete the information requested in a case report form (CRF). Once completed, save and submit the metadata form so that the person analyzing the data could be aware of the context in which the data collection took place.

To access the collected data, follow these steps:

1. After each session with a participant, make sure the data are immediately synchronized in the app on the tablet.

2. Once the data is uploaded, it should be available in the dashboard online. Here, users can interact and view individual data points for each participant. Be aware that, given the sensitivity of this information, the raw data is automatically stored in a cloud-based database that must be specifically created for this purpose.
3. To download the data, one needs to create a dataset on the application dashboard. To do this, select the IDs of the participants you are interested in, and then define the name and description of the dataset.
4. After creating the dataset, a script is generated that can be used to download the dataset. Depending on the tech solutions that have been implemented by your vendor, it is also possible to generate a link that can be used to download the dataset directly from the cloud bucket. A similar path can be followed to download the data from the cloud service of your choice.

Please remember that biomarker data is sensitive and therefore should always be encrypted when downloaded or transferred between individuals, as required by the Institutional Review Board (IRB) that approved your study protocols.





## 7.2 Analyze electrocardiogram data

A standard way to use ECG data in child development research is to examine changes in indicators derived from ECG in response to mild to moderately stressful stimuli. Stress reactivity may be reflected in the difference in ANS indicators from the first (neutral) video to the second (fearful) video. In the Child Development Study in Malawi, participants were outfitted with an ECG device, during which ANS indicators were collected while the participant viewed three 2-minute videos while sitting quietly. The videos differed in stress level in order to allow changes in stress reactivity. In particular, stress regulation/recovery may be reflected in examining the differences in ANS indicators between the second (fearful) video and the third (neutral) video, i.e., regulation following cessation of stressor, and/or between the first (neutral) and third (neutral) videos, i.e., return to baseline.

### **BOX:**

The first and last videos were a clip meant to be neutral (i.e., not elicit a stress response), showing underwater animals swimming while calm music played. The middle video was meant to be slightly frightening (to elicit a stress response), showing large cartoon dinosaurs chasing baby dinosaurs. The videos were presented sequentially without breaks in this order: video neutral, video fearful, video neutral.

Once the ECG data are obtained, the data manager can perform paired-sample t-tests to determine whether there are statistically significant changes at the level of the entire analysis sample in at least four ANS indicators (mean HR, RMSSD, SDNN, stress index; see below) in response to the protocol videos.

It should be noted that the absence of sample-level differences does not mean that there are no individual (or grouped) differences in these scores that might be correlated with or predictive of other developmental variables of interest. Thus, a second dimension of the analysis may investigate, for example, differences in ECG indicators among siblings living in the same household to study how children may differ in relation to specific household dynamics. Below we describe some of the most common indices that are outputted from the MAWI software and could be used in analyses involving ECG data.

- 1.** Mean HR (heart rate) is the mean number of cycles of contraction and subsequent relaxation of the heart, expressed as beats per minute.
- 2.** Stress Index (SI) or SI reflects the extent to which one experienced physiological stress in response to a stressor. In general, a higher stress index is an indicator of higher stress levels. However, a very low stress index ( $< 25$ ) is an indicator of a “rigid heart rhythm,” suggesting exhaustion. Note: Research on the stress index has only been tested in adults, so it is unclear what to expect for the performance or interpretation of this variable in children.
- 3.** Mean NN Interval (NNI) is the mean value of normal intervals RR (NN). The normal-to-normal (NN) interval represents the RR interval data, with artifacts and noise removed, as artifact and noise can make some RR intervals unreliable.
- 4.** SDSD is the standard deviation of successive differences of the RR intervals. SDSD represents short-term variability.
- 5.** SDNN is the standard deviation of the value of normal intervals RR (NN). Abnormal RR intervals are excluded. SDNN is one of the main indicators of heart rate variability (HRV), which reflects the state of regulation mechanisms. SDNN is an integral indicator that characterizes HRV as a whole and depends on the influence of the sympathetic (SNS) and parasympathetic (PNS) divisions of the ANS on the sinus node. An increase or decrease in this indicator shows a shift in the autonomic balance towards the predominance of one of the departments.
- 6.** NNI-20 is the number of pairs of consecutive NN intervals for the episode that differed by more than 20 ms.
- 7.** pNNI20 (ratio of NN20 vs total number of NNI, %) is the percentage of pairs of consecutive normal RR time intervals that differed by more than 20 ms for the episode.
- 8.** Root Mean Square of the Successive Differences (RMSSD) is the square root of the mean of the sum of the squares of differences between adjacent NN intervals. RMSSD is one of the indices that is considered to be a purer marker of PNS activity (as opposed to PNS + SNS), as it indexes cardiac vagal tone. RMSSD reflects vagal tone and is highly correlated with high-frequency HRV.
- 9.** Median-NNI is the median value of the NN interval. The normal-to-normal interval represents the RR interval data, with artifacts and noise removed.
- 10.** Range-NNI is the difference between the largest and smallest values of the NN interval
- 11.** CVSD is the ratio of RMSSD and mean normal-to-normal intervals.
- 12.** CVNNI is the ratio of SDNN and mean normal-to-normal intervals.
- 13.** STD-HR is the standard deviation of instantaneous heart rate values.



### 7.3 Clean outputted data

Although the MAWI software program includes automated processes for managing artifacts, steps should be taken to ensure the cleanest possible dataset following output from MAWI and before data analyses. The following should be considered when deciding whether to retain or remove a participant's data from the final analytic dataset. Note, all decisions regarding data cleaning should be documented and described in any reports.

1. Check the dataset for problematic values (e.g., duplicate participant IDs, incorrect participant IDs). If errors in the dataset cannot be resolved by reviewing the CRF and/or conferring with the study team, remove the affected data.
2. Review the CRF. If there are any notes indicating that the session is invalid (e.g., child did not pay attention to any of the videos, child was very agitated or ill during session), remove the participant's data.
3. You may see abnormal values like "-1" or "NA" in cells in the dataset. These values should be removed, as they do not contain valid data.
4. Check the data for values that are not physiologically realistic. The easiest way to check for such values is to review the mean HR for each episode for each participant. Compare the values to the average HR for a participant of that age and sex/gender. For example, normal resting heart rate for children 3 to 4 years old is approximately 80 to 120 beats per minute, and for children 5 to 6 years old is 75 to 115 beats per minute. Note that this is heart rate during quiet, non-stressed rest. Therefore, HR may be elevated much above these values when stressed and still be considered physiologically valid.
5. Identify outliers in the data. Typically, outliers are defined as values being  $\pm 3$  standard deviations (SD) from the mean for the sample for a given episode. We use mean heart rate to determine outliers. Remove outlier data.
6. If comparison of ECG indicators among protocol episodes is important for your analysis (e.g., comparing response to neutral versus fearful video), you may consider removing participant data if they do not have cleaned data available for all episodes of interest or for at least a minimum number of episodes (e.g., if have first neutral video and fearful video, you will be able to compute stress reactivity but not recovery, and that may be sufficient for your needs to consider retaining).
7. If you are interested only in comparing participant groups (e.g., younger siblings versus older siblings), you may only want to include participants for whom data are available for both types of siblings within a household.
8. After completing all of the data cleaning, review to see if you need to remove any participants due to missing episode data (see point #6 above) and/or group data (see point #7 above).

#### Useful references:

Rudd KL, Roubinov DS, Jones-Mason K, Alkon A, Bush NR. Developmental Consequences of Early Life Stress on Risk for Psychopathology: Longitudinal Associations with Children's Multisystem Physiological Regulation and Executive Functioning. *Dev Psychopathol.* 2021 Dec;33(5):1759-1773. doi: 10.1017/s0954579421000730. Epub 2021 Dec 7. PMID: 34949903; PMCID: PMC8691149.

Rudd KL, Bush NR, Alkon A, Roubinov DS. Identifying profiles of multisystem physiological activity across early childhood: Examining developmental shifts and associations with stress and internalizing problems. *Psychoneuroendocrinology*. 2021 Jun;128:105196. doi: 10.1016/j.psyneuen.2021.105196. Epub 2021 Mar 12. PMID: 33765640; PMCID: PMC8188642.

Ritz T, Schulz SM, Rosenfield D, Wright RJ, Bosquet Enlow M. Cardiac sympathetic activation and parasympathetic withdrawal during psychosocial stress exposure in 6-month-old infants. *Psychophysiology*. 2020 Dec;57(12):e13673. doi: 10.1111/psyp.13673. Epub 2020 Oct 13. PMID: 33048371; PMCID: PMC8548071.

Stephens M, Bush N, Weiss S, Alkon A. Distribution, Stability, and Continuity of Autonomic Nervous System Responsivity at 18- and 36-Months of Age. *Biol Res Nurs*. 2021 Apr;23(2):208-217. doi: 10.1177/1099800420943957. Epub 2020 Jul 27. PMID: 32715727; PMCID: PMC8264862.

Bush NR, Caron ZK, Blackburn KS, Alkon A. Measuring Cardiac Autonomic Nervous System (ANS) Activity in Toddlers - Resting and Developmental Challenges. *J Vis Exp*. 2016 Feb 25;(108):53652. doi: 10.3791/53652. PMID: 26967045; PMCID: PMC4828199.

Fortunato CK, Gatzke-Kopp LM, Ram N. Associations between respiratory sinus arrhythmia reactivity and internalizing and externalizing symptoms are emotion specific. *Cogn Affect Behav Neurosci*. 2013 Jun;13(2):238-51. doi: 10.3758/s13415-012-0136-4. PMID: 23233122; PMCID: PMC4464793.



---

# Cognitive development



**Authors:**

Daniel Robles,  
Nicolò Tomaselli, and Guilherme Lichand

## 8. Cognitive development

### **DISCLAIMER:**

This user guide describes the collection of EEG data using a headband device that we will refer to as Muse device. These electrodes are manufactured by the company Interaxon (<https://choosemuse.com/>) and are used to collect EEG data. Because contracting with a vendor for this type of activity plays an important role in successful data collection, we clarify that we are referring to this specific data collection. Other devices could be used, for which the procedures might be different.

Another biomarker increasingly used in daily life is the EEG, or electroencephalogram. This is a technique used to measure and record the electrical activity of the brain. EEG used to be quite invasive, but recent developments in the field are making this procedure available on the shelves for those who want, for example, to improve their sleep or meditative activities.

In practice, electroencephalography involves placing small electrodes on the scalp that detect and record electrical signals produced by neurons in the brain. The usual method of measuring brain waves is to wear a helmet with several sensors and wires on the scalp. This method is considered impractical for collecting data on large samples and under difficult conditions in the field. Another much simpler method of collecting EEG information is to use wireless headphones like Muse that, through a Bluetooth connection and an automated pipeline, process EEG data and generate indices that can be used for data analysis and comparison. The following parameters can be measured.

- 1. Brain waves.** There are different types of brain waves that are associated with different states of consciousness, such as alpha waves during relaxation and theta waves during deep sleep. Alpha waves are most commonly observed when a person is awake but relaxed, with eyes closed. They are often associated with a state of calmness, relaxation and reduced mental effort.
- 2. Amplitude.** The amplitude of brain waves represents the strength of electrical activity in the brain. A greater amplitude usually indicates more intense brain activity.
- 3. Frequency.** The frequency of brain waves represents the speed at which the waves oscillate. Different frequencies are associated with different states of consciousness, such as beta waves during focused attention and delta waves during deep sleep.
- 4. Coherence.** It is a measure of the degree to which brain waves are synchronized between different regions of the brain. Greater coherence may indicate greater connectivity and communication between brain regions.

5. Event-related potentials (ERPs). ERPs are electrical signals that occur in response to specific stimuli, such as a flashing light or sound. These signals can provide insight into cognitive processes such as perception, attention and memory.
6. Power spectral density: Power spectral density is a measure of the power distribution among different brainwave frequencies. This metric can provide information about the overall balance of brain activity and the dominant frequency bands in a particular state of consciousness or cognitive process.

## 8.1 Collect electroencephalogram data

A detailed protocol is necessary to collect accurate EEG data and ensure its consistency and reliability. The procedures and protocols described here aim to provide accurate and complete instructions for EEG data collection, including electrode placement, recording techniques, and basic interpretation of results.

This user-guide describes EEG data collection supported by a headset device produced by the company Interaxon (<https://choosemuse.com/>). Specifically, we refer to the use of the Muse2 device, which we will call Muse headband for simplicity. The Muse headband has seven sensors that detect brain activity, including five on the forehead and two behind the ears. Optionally, one auxiliary electrode can be simply installed to collect more data. An auxiliary electrode is an additional electrode used to help improve signal quality. In fact, EEG measures electrical activity in the brain using electrodes placed on the scalp, and the signal can be affected by various factors, such as electrode placement, skin impedance and interference from other electrical sources.

Muse headband uses Bluetooth technology to transmit data to an app installed on a tablet device. Brain wave data can be collected using these devices which can be found at retailers or purchased online. The device is placed on the child's head, with electrodes in contact with the skin, for five minutes. The band is adjustable to fit the size of the head and for younger children can be secured with a piece of cloth. The following procedure, while specific to Muse headband, could easily be applied to any other device capable of monitoring this biomarker.

Before starting, the researcher should make sure to choose a quiet and comfortable place to begin the session and to have these items available.

- Charged tablet
- Conductive Gel
- Cotton swab/wet wipes
- Stretch gauze
- Charged Muse headband
- Setting up the tablet

To collect the EEG data, the following steps should be undertaken:

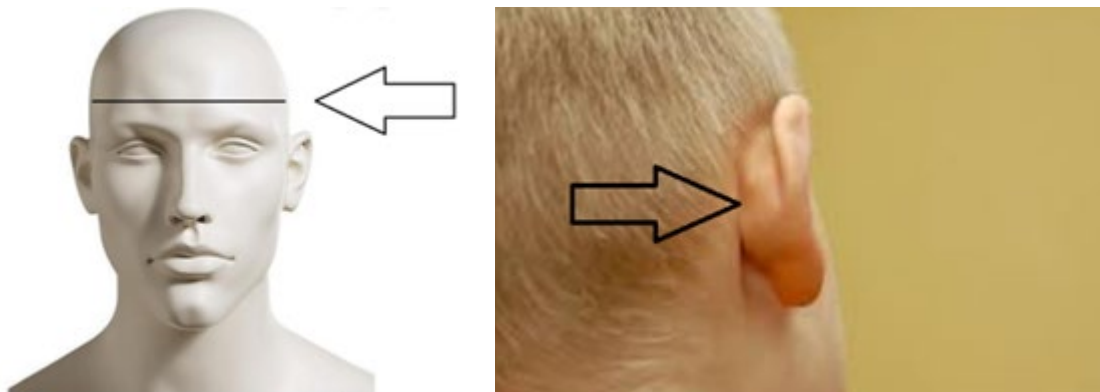
1. Make sure the tablet and Muse headband are charged. Note that the Muse headband requires a software application to transmit data to the tablet. In the procedure reported here, the same software application is used for ECG and EEG data. Make sure that the latest version of the application is available. If the latest version of the application is not available on the tablet, be sure to download it or contact the vendor.

2. Click on the application on the tablet, open it and select the work area. Here you enter the data for: Region, District and village name. Select the family and fill in all the required details, including informed consent and family information. Next, guardian and child information should be entered.
3. The Muse headband is connected to the tablet via Bluetooth. Connect the Muse headband to the tablet by clicking on the same Muse device ID. Once this is done, return to the participant profile and choose the name of the participant from whom you want to collect data.
4. To set the Muse headband on the child, use an alcohol swab to wipe the forehead and the area behind the participant's ear. If you use the auxiliary electrode, clean the center of the child's head where the auxiliary electrode is in contact with the skin. Then connect the "aux" cable into the Muse. The auxiliary port is located on the right side of the Muse device.
5. Then, with the help of a cotton swab, insert it into the gel container to get enough gel to apply to each of Muse's 5 front sensors and the two rubber ear sensors, as well as the auxiliary metal side.
6. Place the muse band directly above the eyebrows. The muse band should sit above your finger. Ensure that the Muse is centered on the in alignment with the participant's nose. Ensure that the sensors behind the ear are sitting evenly.
7. While standing behind the person, adjust the muse length so sensors behind the ear make contact with the back of the ears. If the child's head is too small for the Muse to fit properly, use an elastic gauze piece to help the Muse make better contact with the child's skin (see picture below).
8. If you use the auxiliary electrode please notice the "Pz" area shown and circled in the picture below. Note, you must secure the auxiliary electrode with another piece of elastic gauze or with a stretchy material that is comfortable and not abrasive to the child. By holding the electrode in place we ensure that the auxiliary connectivity remains high quality when the child moves.
9. Use the same elastic gauze to also hold the auxiliary electrode in place as shown below. Please make sure the auxiliary electrode is placed under the elastic gauze (as shown below) to minimize the cable movement during the data collection.

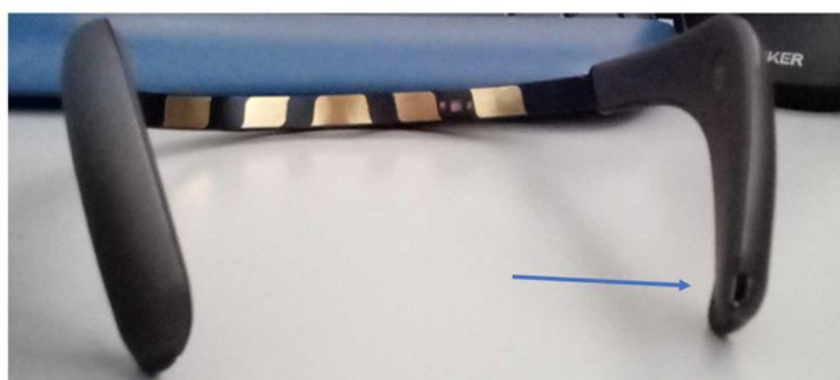
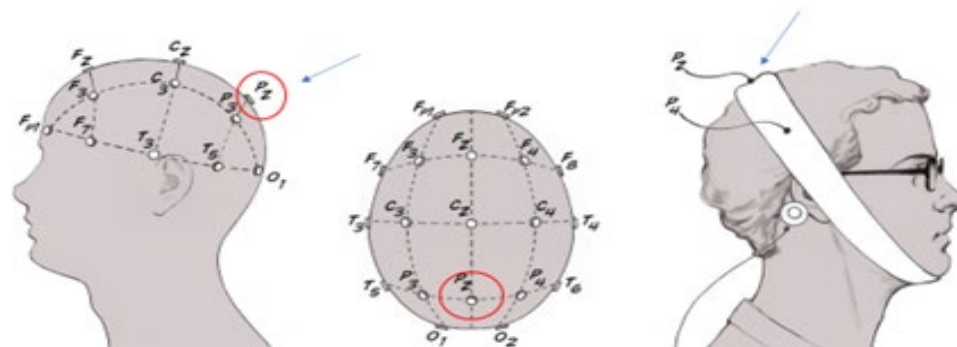
Next, record for 5 minutes of resting state data while the participant has fully closed their eyes and make sure that they stay very still until the end of the session. Ensure that the child is either looking at the tablet while the data is recording or that the child is looking at a location without a lot of activity. So you might want the child to be facing a wall instead of a busy road, or a busy window with noises and things moving. Remember that the less movement, noises and distractions, the better the data quality.

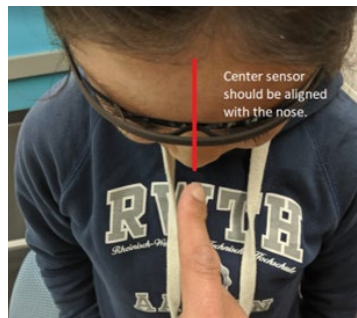
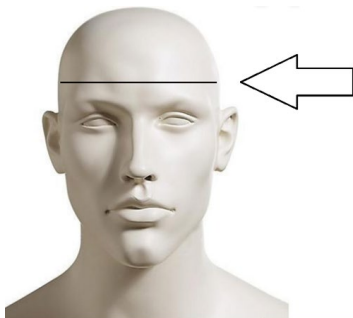
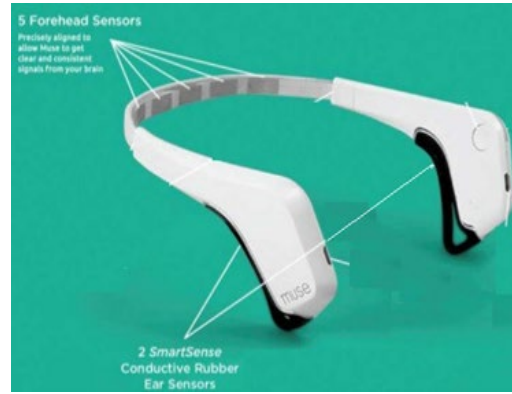
To check if the data recorded is satisfying the required minimum quality, watch the graph visualization of the recording on the tablet, see sample images below. After recording, clean the muse and electrodes gently using an alcohol swab or pad. Also clean the gel residue from the child's skin. In order to access the data collected, make sure that after every session with a participant, the data is synced in the Mawi application immediately. Once the data is uploaded, it can be accessed in the Mawi application dashboard. Here, users can interact with and visualize individual data points for each participant. The raw data is stored in a cloud-based bucket that is specifically created for this purpose.

In order to download the data, you need to create a dataset on the app dashboard. To do this, select the participant IDs that you are interested in and then set the dataset name and description. After creating the dataset, a script is generated that can be used to download the dataset. Alternatively, a link is also generated that can be used to download the dataset directly from the cloud bucket. Similar path could be followed to download data from the cloud service of your preference. Please remember that bio-marker data is sensitive and as such it should always be encrypted when downloaded or transferred between individuals, according to the prescriptions of the Institutional Review Board that approved your study protocols.

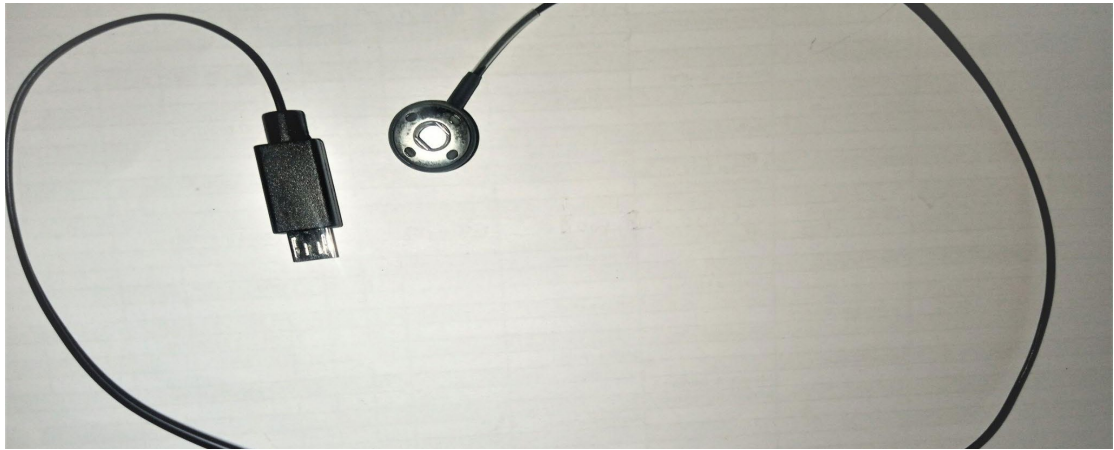


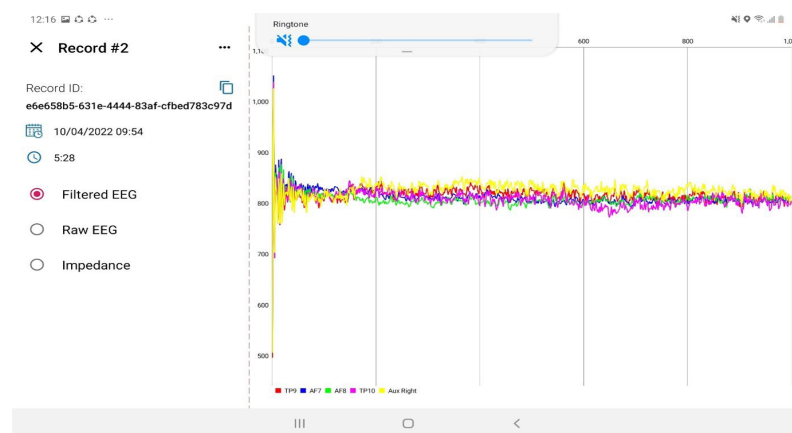
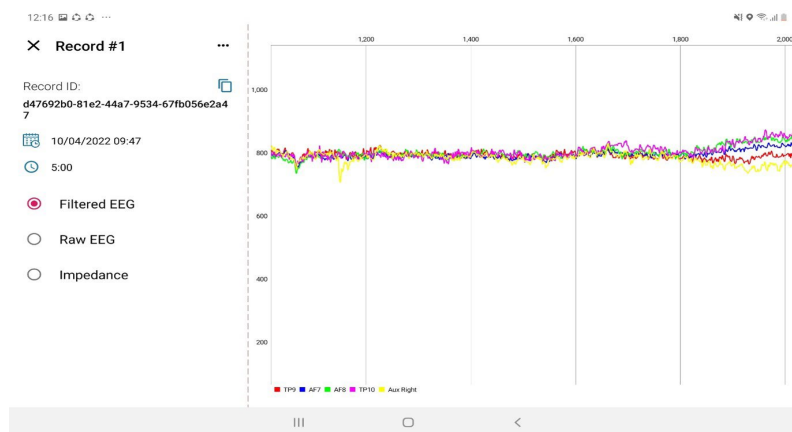
## Location of aux electrode. Pz area











## 8.2 Analyze electroencephalogram data

Several studies have suggested that in the early years of life, children may have different power in the low-frequency bands of the EEG (i.e., theta) or in the mid- and high-frequency bands (i.e., alpha, beta, and gamma). These differences are correlated with several sociodemographic factors. Of course, these general patterns conceal considerable heterogeneity; not all children in similar situations will show these neurodevelopmental differences. However, this analysis proves to be relatively simple and rich when combined with other data streams.

What follows are specific and technical instructions on how to manipulate EEG signals. While recognizing that this is a rather specific task requiring ad hoc expertise, we provide what we believe to be useful pointers for experts who will exploit the data. The analysis of EEG signals that we report here is based on a specific setup supported by software called BrainVision Analyzer. We report here the link available to this proprietary software at the time of writing this user guide: <https://brainvision.com/products/analyzer-2/>.

To study resting state spectral analysis for EEG data using the BrainVision analyzer, it is first necessary to convert Muse output files to BrainVision analyzer files. A script, based on Matlab, that can be used for this purpose is available at this link.

[https://github.com/d-robles/Convert\\_Muse\\_Files/tree/master/Convert%20Muse%20PEER](https://github.com/d-robles/Convert_Muse_Files/tree/master/Convert%20Muse%20PEER)

Once successfully converted, the data are filtered using IIR filters with a bandpass filter range of [0.1 30] Hz, with a notch frequency set at 60 Hz. The data are segmented into epochs using the [-200 800] ms interval. The data are then baseline corrected using an interval of [-200 0] ms based on the segmented epochs.

**BOX:** Artifact Rejection

In the pop-up window, select Automatic Segment Selection on the Inspection Method tab. Uncheck all other options on this tab.

Switch to the Channels tab. Select the Enable all option.

Switch to the Criteria tab, and on the Gradient subtab, select Check Gradient and enter 10 in the Maximum Allowed Voltage text box. In the Before Event space, enter 200. Do the same in the After Event text box.

Now switch to the Min-Max sub tab and select Maximum Allowable Absolute Difference. In the Next text box, enter 100.

In the Interval Length text box, enter the length of the segment (e.g., 800 ms: see above). In the Before event text box, enter 200. Do the same in the After event text box. Finally, terminate the artifact rejection by clicking OK. Report the data to the average of electrodes TP9 and TP10.

Once these preliminary steps have been completed, the researcher can begin to perform rapid Fourier transform (FFT) analysis using the following parameters.

---

\*\*\* Fast Fourier Transformation (FFT) \*\*\*

*Maximum Resolution*

---

*Power*

*Non-Complex Output*

*Half Spectrum used*

---

*Data Window:*

*Hanning Window*

*Length = 10 %*

*Variance Correction not used*

*Periodic*

---

*Segments have been normalized*

*Selected range: 0.5 Hz - 30 Hz*

---

\*\*\* Data node specific information \*\*\*

*Resolution: 1 Hz*

*Computed normalization range: 1 Hz - 30 Hz*

---

## Useful references:

- Alahmadi, N., Evdokimov, S. A., Kropotov, Y. J., Müller, A. M., & Jäncke, L. (2016). Different resting state EEG features in children from Switzerland and Saudi Arabia. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2016.00559>
- Boersma, M., Smit, D. J. A., Bie, H. M. A. de, Baal, G. C. M. V., Boomsma, D. I., Geus, E. J. C. de, Waal, H. A. D. de, & Stam, C. J. (2011). Network analysis of resting state EEG in the developing young brain: Structure comes with maturation. *Human Brain Mapping*, 32(3), 413–425. <https://doi.org/10.1002/HBM.21030>
- Clarke, A. R., Barry, R. J., Dupuy, F. E., Heckel, L. D., McCarthy, R., Selikowitz, M., & Johnstone, S. J. (2011). Behavioural differences between EEG-defined subgroups of children with Attention-Deficit/Hyperactivity Disorder. *Clinical Neurophysiology*, 122, 1333–1341. <https://doi.org/10.1016/j.clinph.2010.12.038>
- Giertuga, K., Zakrzewska, M. Z., Bielecki, M., Racicka-Pawlukiewicz, E., Kossut, M., & Cybulska-Klosowicz, A. (2017). Age-related changes in resting-state eeg activity in attention deficit/hyperactivity disorder: A cross-sectional study. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2017.00285>
- Katada, A., Ozaki, H., Suzuki, H., & Suhara, K. (1981). Developmental characteristics of normal and mentally retarded children's EEGs. *Electroencephalography and Clinical Neurophysiology*, 52(2), 192–201. [https://doi.org/10.1016/0013-4694\(81\)90166-8](https://doi.org/10.1016/0013-4694(81)90166-8)
- Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., & Colino, F. L. (2017). Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in Neuroscience*. <https://doi.org/10.3389/fnins.2017.00109>
- Louis, E. S., Frey, L., Britton, J., Hopp, J., Korb, P., Koubeissi, M., Lievens, W., & Pestana-Knight, E. (2016). Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants. In *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. <https://doi.org/10.5698/978-0-9979756-0-4>
- McEvoy, K., Hasenstab, K., Senturk, D., Sanders, A., & Jeste, S. S. (2015). Physiologic artifacts in resting state oscillations in young children: Methodological considerations for noisy data. *Brain Imaging and Behavior*, 9(1). <https://doi.org/10.1007/s11682-014-9343-7>



# Metadata



**Authors:**

Michelle Bosquet Enlow,  
Nicolò Tomaselli, and Guilherme Lichand

## 9. Metadata

A crucial aspect of any survey is the need to collect information about the context in which the data were collected. This information is usually called metadata. In other words, it is necessary to provide data about other data. Not about their content, but about the setting in which they were collected. Was the data collected at the participant's home or at school? Was the child's guardian present? Did the child seem stressed or nervous about the devices? What are the reasons why some data were not collected?

Depending on the type of data you are dealing with, metadata can be captured by different methods. There are several types of metadata, including the following.

1. **Descriptive metadata.** It is also called process data, and it describes the processes of collecting, processing or producing statistical data and the devices used (tablets, pads, sensors, GPS, etc.).
2. **Structural metadata.** It provides information about data architecture and the organization of information.
3. **Administrative metadata.** It provides complementary information from other sources and how these data sources may be used.
4. **Legal metadata:** include information about informed consent or assent required/obtained from participants.

As an example, here below is the Metadata collected in the context of the Child Development Study in Malawi to describe the cardiac ECG data collection. This form was specifically called Care Report Form or (CRF).

### **\*\*EXAMPLE OF METADATA FORM\*\***

CDS: ANS VISIT CASE REPORT FORM

Date \_\_\_\_\_

Primary Enumerator \_\_\_\_\_

Secondary Enumerator \_\_\_\_\_

Time of Arrival to Home/Clinic \_\_\_\_\_

#### CHILD ANTHROPOMETRIC MEASURES

Height (cm) \_\_\_\_\_ Weight (kg): \_\_\_\_\_ Age (in days): \_\_\_\_\_

Sex: Male/Female

#### EQUIPMENT USED

BioRadio    MAWI

Rate the child's amount of movement during the protocol: \_\_\_\_\_

1. Sat quietly for most/all of protocol.
2. Moderate movement for portion of protocol.
3. Frequent movement within the seat or getting in and out of the seat during much of protocol.

Rate the child's compliance with instructions: \_\_\_\_\_

1. Followed instructions for all of the protocols.
2. Needed to be reminded a few times to follow instructions (e.g., to sit still, not talk, keep fingers on MAWI pad).
3. Needed to be repeatedly reminded to follow instructions (e.g., to sit still, not talk, keep fingers on MAWI pad).

Rate the child's attention to the stimuli: \_\_\_\_\_

1. Paid attention to stories and videos throughout protocol.
2. Needed to be reminded a few times to pay attention to protocol. Brief periods of not paying attention.
  - a. Specify episodes when child did not pay attention: \_\_\_\_\_
3. Needed repeated reminders to pay attention to protocol. Significant time not paying attention.
  - a. Specify episodes when child did not pay attention: \_\_\_\_\_

Any Part of Protocol Refused?                      No     /     Yes Which Part(s)  
\_\_\_\_\_

Protocol Ended Early?                      No     /     Yes

General Notes:

Before beginning any data collection with human subjects, informed consent must be obtained. Informed consent can be verbal or written, and it is one of the basic principles of research ethics. Its purpose is for human participants to enter research freely (voluntarily) with full information about what it means for them to participate, to give consent before entering the research, and to be able to withdraw from it as soon as they wish.

It is good practice to collect other metadata for the various data streams through the use of computer assisted personal interviewing (CAPI) tools. In other words, small and agile survey tools should be designed to collect that information that cannot be easily captured otherwise. The data manager should be careful to store this information in encrypted folders, since metadata is oftentimes a personal identification information (PII).

Finally, it should be noted that administrative data can also be collected and is a source of information that could be merged with the rest of the data. Administrative data can sometimes be collected by governments or other local institutions or organizations for monitoring purposes. When used in conjunction with firsthand data, they can help to assess the context in which the study takes place. Birth records or school enrollments are examples of administrative data. These types of data can be used to produce important information in a cost-effective way. When turned into indicators, they can show trends over time and reflect the actual context.



# 10

---

Data quality  
check-lists and  
must dos

**Authors:**

John Phuka, Beverly Laher, Lonjezo Dzekedzeke,  
Nicolò Tomaselli, and Guilherme Lichand

## 10. Data quality check-lists and must dos

Preliminary data checks are an essential part of any data collection process. These checks allow researchers to ensure that the data used are reliable, accurate, and complete before beginning analysis. The purpose of these checks is to identify any errors or inconsistencies in the data that could affect the results of the analysis. By conducting these checks, researchers can ensure that they are drawing valid conclusions based on sound data. Preliminary data checks are therefore a crucial step in the research process, as they help ensure the reliability and replicability of the results.

- 1. Survey data.** In recent years, the quality of data collected face-to-face has improved dramatically through the use of Computer Assisted Personal Interviewing (CAPI). Software available in the field allows for many integrated schemes that aim to reduce enumerator errors and increase data quality. However, some concepts are still useful for successful data collection: be side-by-side of enumerators, perform random spot checks to verify operations, run audits or follow-up checks, and use a missing-persons form to track missing participants. These are the main guiding principles not to be forgotten.
- 2. Anthropometric.** The most common errors in anthropometry are body positioning, measurement reading, and recording. To minimize these errors, standard procedures for obtaining measurements should be adopted, such as: proper training, automated data entry systems, and procedures for paper measurement forms. In addition, equipment should be calibrated frequently, and examiners and recorders should be observed periodically by more experienced and qualified examiners to ensure standardization. All equipment in the body measurement room should be checked, maintained, and cleaned regularly to protect it. All equipment that is broken or starting to break should not be used. Ensure that the child does not suffer any physical harm or risk while performing the measurements. Children should be monitored and given additional safety precautions during body measurement. Children in the body measurement room need continuous attention from technicians. Children should be carefully held by the technician when using the body measurement table or supine length table to avoid accidents. Remember that toddlers and infants often turn over very quickly. To avoid the risk of falling objects on a child using the recliner table, mount equipment baskets on the wall.
- 3. Language and speech development.** Collecting high-quality audio recordings necessary for accurate analysis and meaningful insights. To ensure the quality of the collected audio data, several data quality checks could be performed. After downloading the recordings from the recorders, it's important to check that the recorded files are audible and that they are played back within the allotted period of time, which is typically 10 minutes to 12 hours depending on the study design. Any recordings that are inaudible or not played back within the stipulated time may need to be excluded from the analysis. In addition, enumerators should be trained to indicate at the beginning

of each session the participant ID, the current recording episode, and the child being recorded. This information helps to ensure that the recorded data can be accurately attributed to the correct participant and recording episode. Enumerators must also record with their own voice to maintain consistency across recording sessions. Sometimes, recordings may be split into multiple episodes because some participants may tamper with the recorders. In such cases, recorders have a built-in application that can be installed in a computer and used to combine the files into one longer file. This helps to maintain the integrity of the data and ensures that the recordings can be accurately analyzed. To protect against data loss, it's recommended to create an encrypted backup of the audio recordings by storing copies in an external hard drive or computer. This helps to ensure that the data can be easily retrieved in case of any unforeseen events such as data corruption or loss. Overall, performing these data quality checks is crucial to ensure that the collected audio data is of high quality and suitable for analysis.

4. **Social Interactions.** The first step in checking the quality of the proximity sensor data is to download and inspect it. Once downloaded, the application will display an interface that indicates the number of bits that are contained in the sensor. These bits represent the memory capacity that has been filled by the data being collected by the sensors when it has been in contact with other sensors. Each contact recorded between the sensors is represented as one bit. If a sensor contains less than 20 bits after being worn by a participant for a period of time, it's highly likely that the sensor does not contain quantifiable data and should be excluded from the analysis. Make sure that you enter the sensor barcode in the web app and is mapped with the sensor hardware ID displayed in the web app. This mapping will be captured inside a JSON file that can be downloaded and the mapping structure can be analyzed using a text editor. Check for duplicates inside the JSON file and redundancies that may occur by entering an incorrect sensor ID number. In addition, it's important to ensure that metadata containing the corresponding sensor ID mapped to the hardware ID is generated for each sensor from which data is downloaded. This metadata will help to identify which sensor was used to collect the data, and ensure that the data can be accurately attributed to the correct sensor. Lastly, before uploading the metadata file to the cloud server, it's essential to check that the file is not corrupted. TIP: If downloading data from more than a couple of hundred sensors, a good approach is to download data in batches of less than 50, as errors in the acquired sensors can be tracked and resolved more easily.
5. **ECG.** Several data quality inspections ought to be carried out to guarantee the accuracy of the recorded ECG data. Primarily, it's crucial to confirm that the data is accurate and that no values, parameters, or columns are missing. Verify that all data files and attributes have been included in each participant's cloud-based data. The accuracy and validity of the study could potentially be undermined by any missing values or incomplete data. The dashboard must be inspected to confirm that the recorded data can be viewed and is consistent within and across all parameters before confirming data consistency. Data inconsistencies may result in flawed analysis and inaccurate conclusions. Also, it's important to look for outliers in the data, or values that differ markedly from the rest of the data. It is critical to confirm whether outliers are real data points since they may refer to errors in measurement or data entry. They ought to be kept in the analysis if they are reliable data items. But if they are invalid, they ought to be eliminated from the dataset. Duplicates, which can happen when a researcher administers a test to a person twice and obtains two data points for that individual, must also be examined. As a general rule, it is advised to perform data analysis during the initial test session in order to prevent duplicate values or files.

6. EEG. One of the first steps in checking the quality of EEG data is to ensure that the data is complete and there are no missing values, variables, or columns. The downloaded data from Muse should contain two files for each session, typically one csv and one json. These files should contain all the necessary data, including timestamps, participant IDs, and EEG measurements. Any missing data or incomplete files could compromise the accuracy of the analysis and lead to incorrect conclusions. Furthermore, it's important to check the data consistency both within and across all variables. This can be done by verifying on the dashboard that the recorded data can be viewed and that it is consistent. Any inconsistencies could also compromise the accuracy of the analysis and lead to incorrect conclusions. Another important quality check for EEG data is to look for outliers and absurd data. Identifying and correcting these outliers is crucial to ensuring that the data analysis is accurate and reliable.

Below are three checklists that could be used as an example of a management tool to ensure that the field team has everything under control, before, during, and after data collection. The reader should consider these tables as templates to be adapted to the specific context and data collection effort.

**Table A – Data Collection Plan, from source to server**

Markers	Source of data	Interface to collect data	Interface to retrieve data	Participant-ID	Server	Sample
<b>MAWI</b>	ECG data for each child	Application developed by MAWI and running in a tablet	Tablet synchronized to online dashboard	Entered manually on Mawi App when creating participant's profile	AWS plus dashboard interface Automatic synchronization while tablet online	300 children
<b>MUSE/ INTERAXON</b>	EEG data for each child	Headband wore by child				
<b>ISI</b>	Face-to-face contact among participants within/among households	Metadata collected by Jotform and running in a tablet	Laptop accessing a webapp for data download	Preload on Jotform, enumerator will select from list	Files uploaded on AWS. Bucket Proximity Sensors data created and shared with Ciro. Lonjezo upload through AWS CLI and Ciro download through AWS CL	600 children plus 300 guardians
<b>VOICE RECORDER</b>	Speech and language	Recorders wore by children	Laptop running usual suite	Preload on Jotform, enumerator will select from list. Plus, file naming once extracted	Files manually uploaded on AWS. Bucket for Verbal Interactions data is created	300 children

Last update: DATE

Table B – Checklist for fieldwork data collection

Protocol	Face-to face	Check	Wearable technology
<b>MAWI</b>	Mawi Native form in Mawi App		
	Is it installed on every tablet?	<input type="checkbox"/>	Clamps are working? <input type="checkbox"/>
	It works correctly?	<input type="checkbox"/>	Chargers are sufficient? <input type="checkbox"/>
<b>MUSE/ INTERAXON</b>	Jotform ECG case report		Mawi Pads (7)
	Is it installed on every tablet?	<input type="checkbox"/>	Dust protection bags? <input type="checkbox"/>
	It works correctly?	<input type="checkbox"/>	Auxiliary electrodes? <input type="checkbox"/>
			Gel is sufficient? <input type="checkbox"/>
			Dust protection bags? <input type="checkbox"/>
<b>ISI</b>	Jotform Proximity Sensors		Muse Headbands (30)
	Is it installed on every tablet?	<input type="checkbox"/>	Wires and chargers? <input type="checkbox"/>
	It works correctly?	<input type="checkbox"/>	Two programmers? <input type="checkbox"/>
			Batteries are sufficient? <input type="checkbox"/>
			Two timestamps fire starters? <input type="checkbox"/>
<b>VOICE RECORDER</b>	Jotform Language interactions		Open beacon tags (950)
	Is it installed on every tablet?	<input type="checkbox"/>	Configuration is OK? <input type="checkbox"/>
	It works correctly?	<input type="checkbox"/>	Batteries are sufficient? <input type="checkbox"/>
			Dust protection bags? <input type="checkbox"/>
			Tablets are in sufficient number and working? <input type="checkbox"/>
<b>TRACKING BASELINE</b>	Survey CTO baseline		Wires and chargers? <input type="checkbox"/>
	Is it installed on every tablet?	<input type="checkbox"/>	Laptops for data retrieve and backstopping? <input type="checkbox"/>
	It works correctly?	<input type="checkbox"/>	Enumerators Bags <input type="checkbox"/>
	Survey CTO IDELA/MDAT		Did you check to have spares of everything? <input type="checkbox"/>
	Is it installed on every tablet?	<input type="checkbox"/>	Other? <input type="checkbox"/>
	It works correctly?	<input type="checkbox"/>	
			Last update DATE

**Table C – Checklist for fieldwork before / during / after**

Data Stream	Checklist Items	Before the field	During the field	After the field	
ECG	Tablet	<ul style="list-style-type: none"> <li>- Mawi app installed</li> <li>- Jotform installed</li> <li>- Internet bundle active</li> </ul>	<ul style="list-style-type: none"> <li>- Charged</li> <li>- Synchronize after every test in the Mawi app.</li> <li>- Jotform CRF form should be filled for every session.</li> </ul>	<ul style="list-style-type: none"> <li>- Check in the Mawi dashboard if all records are reflecting.</li> <li>- Check in the Jotform dashboard if records are reflecting.</li> <li>- Check in all tablets that all records are synced.</li> <li>- Store in locked cabinet</li> </ul>	
	Mawi pad	<ul style="list-style-type: none"> <li>- Charged and working</li> </ul>	<ul style="list-style-type: none"> <li>- Charged before every session</li> </ul>	<ul style="list-style-type: none"> <li>- Clean and store in dust protection bags</li> </ul>	
	Kitchen VEG board	<ul style="list-style-type: none"> <li>- Check condition</li> </ul>	<ul style="list-style-type: none"> <li>- Carry for every session</li> </ul>	<ul style="list-style-type: none"> <li>- Clean and store</li> </ul>	
	A chair	<ul style="list-style-type: none"> <li>- Check condition</li> </ul>	<ul style="list-style-type: none"> <li>- Carry for every session</li> </ul>	<ul style="list-style-type: none"> <li>- Clean and store</li> </ul>	
	Power bank and Charger	<ul style="list-style-type: none"> <li>- Charged and working</li> </ul>	<ul style="list-style-type: none"> <li>- Use for charging Mawi pad</li> </ul>	<ul style="list-style-type: none"> <li>- Clean and store</li> </ul>	
	Wrist bands	<ul style="list-style-type: none"> <li>- Check condition</li> </ul>	<ul style="list-style-type: none"> <li>- Are used and cleaned after every session.</li> </ul>	<ul style="list-style-type: none"> <li>- Clean and store</li> </ul>	
	Electrodes	<ul style="list-style-type: none"> <li>- Check condition</li> </ul>	<ul style="list-style-type: none"> <li>- Are used and cleaned after every session.</li> </ul>	<ul style="list-style-type: none"> <li>- Clean and store</li> </ul>	
	EEG	Tablet	<ul style="list-style-type: none"> <li>- Mawi app installed</li> <li>- Jotform installed</li> <li>- Internet bundle active</li> </ul>	<ul style="list-style-type: none"> <li>- Charged</li> <li>- Synchronize after every test in the Mawi app.</li> <li>- Jotform CRF form is filled for every participant.</li> </ul>	<ul style="list-style-type: none"> <li>- Check in the Mawi dashboard if all records are reflecting.</li> <li>- Check in the Jotform dashboard if records are reflecting.</li> <li>- Check in all tablets that all records are synced.</li> <li>- Store in a locked cabinet</li> </ul>
		Muse headband	<ul style="list-style-type: none"> <li>- Test charging and connection quality with Mawi app through bluetooth</li> </ul>	<ul style="list-style-type: none"> <li>- Charged and cleaned after every session</li> </ul>	<ul style="list-style-type: none"> <li>- Clean and wrap in bubble wrap.</li> <li>- store in a locked cabinet.</li> </ul>

Data Stream	Checklist Items	Before the field	During the field	After the field
	Conductive Gel	<ul style="list-style-type: none"> <li>- Check expiry date and quantity.</li> <li>- If gel is in a large container, pour in smaller portable containers.</li> </ul>	<ul style="list-style-type: none"> <li>- Carry and use for every session.</li> <li>- Refill every day if being used in a continuous period.</li> </ul>	<ul style="list-style-type: none"> <li>- Store below room temperature.</li> </ul>
	Stretch gauze	<ul style="list-style-type: none"> <li>- Check quality, quantity and condition.</li> </ul>	<ul style="list-style-type: none"> <li>- Carry and use for every session.</li> </ul>	<ul style="list-style-type: none"> <li>- Store in a locked cabinet.</li> </ul>
	Cotton swab/wet wipes	<ul style="list-style-type: none"> <li>- Check quality, quantity and condition.</li> </ul>	<ul style="list-style-type: none"> <li>- Carry and use for every session.</li> </ul>	<ul style="list-style-type: none"> <li>- Store in a locked cabinet.</li> </ul>
Language Interactions	Recorder	<ul style="list-style-type: none"> <li>- Test recorder condition.</li> <li>- Configure the recorder settings if new.</li> </ul>	<ul style="list-style-type: none"> <li>- Configure the date and time before each session.</li> <li>- Check the battery life before each session.</li> <li>- Check the recorder memory. If it contains data from previous sessions, download and Erase the data.</li> </ul>	<ul style="list-style-type: none"> <li>- Check recorder if there is any data and download.</li> <li>- Save the data in folders on a Laptop.</li> <li>- Upload the recordings to the Aws server.</li> <li>- Remove batteries and store them in a locked cabinet.</li> </ul>
	Batteries	<ul style="list-style-type: none"> <li>- Check the quality, quantity and lifespan of the batteries to be used.</li> <li>- This can be done by doing an in-house test by placing the batteries in a recorder and determining how long they last.</li> </ul>	<ul style="list-style-type: none"> <li>- Carry and use for every session.</li> <li>- Check the battery life before each session.</li> <li>- If batteries have less than a half-life, do not use them for long hours recording.</li> </ul>	<ul style="list-style-type: none"> <li>- Dispose the batteries following the recommended waste management practices.</li> </ul>
	Cling wrap	<ul style="list-style-type: none"> <li>- Check quality and quantity.</li> </ul>	<ul style="list-style-type: none"> <li>- Carry and use for every recording session</li> </ul>	<ul style="list-style-type: none"> <li>- Dispose the cling wrap following the recommended waste management practices.</li> </ul>



Data Stream	Checklist Items	Before the field	During the field	After the field
	Vests	<ul style="list-style-type: none"> <li>- Check the quantity and quality.</li> <li>- Ensure there's a variation of sizes to fit all ages.</li> </ul>	<ul style="list-style-type: none"> <li>- Identify the right size to use per session.</li> <li>- Ensure it is clean.</li> <li>- If not clean, please wash with anti-fungal/bacterial soap.</li> </ul>	<ul style="list-style-type: none"> <li>- Clean and store</li> </ul>
	Rubber stopper	<ul style="list-style-type: none"> <li>- Check effectiveness in keeping the recorder on by in-house testing.</li> </ul>	<ul style="list-style-type: none"> <li>- Carry and use for every session.</li> </ul>	<ul style="list-style-type: none"> <li>- Dispose following the recommended waste management practices.</li> </ul>
	Timer/Stopwatch	<ul style="list-style-type: none"> <li>- Set time and date.</li> </ul>	<ul style="list-style-type: none"> <li>- Set time and date</li> <li>- Use for every session(10 minute recording)</li> </ul>	<ul style="list-style-type: none"> <li>- Store in a locked cabinet.</li> </ul>
	Tablet	<ul style="list-style-type: none"> <li>- Install jotform</li> <li>- Activate internet bundle.</li> </ul>	<ul style="list-style-type: none"> <li>- Charge and use every session.</li> <li>- Record on jotform after every session (10 minute session &amp; 12 hours session).</li> </ul>	<ul style="list-style-type: none"> <li>- Store in a locked cabinet.</li> </ul>
Social Interactions	<ul style="list-style-type: none"> <li>- Open beacon sensors</li> </ul>	<ul style="list-style-type: none"> <li>- Check quantity to be used.</li> <li>- Check sensors are programmed.</li> <li>- If they had data from previous sessions, Erase.</li> </ul>	<ul style="list-style-type: none"> <li>- Ensure you have a master sensor with the correct time set to be used for propagating the timestamp.</li> <li>- Make sure the sensors are displaying the correct blinking patterns.</li> <li>- Insert battery during deployment and remove the battery after deployment.</li> <li>- Sensors should be kept in their plastic casings to avoid physical contact.</li> </ul>	<ul style="list-style-type: none"> <li>- Download data from the sensors using the programmer.</li> <li>- Map the hardware ID with the sensor ID.</li> <li>- After downloading, upload the sensor data with the metadata files to the Aws server.</li> <li>- Store the sensors in a locked cabinet.</li> </ul>
	- Sensor casing	<ul style="list-style-type: none"> <li>- Make sure the number of sensor casings matches the number of sensors to be used by physically counting.</li> </ul>	<ul style="list-style-type: none"> <li>- Carry pairs appropriate for use for every deployment.</li> </ul>	<ul style="list-style-type: none"> <li>- Store the casings in a locked cabinet.</li> </ul>

Data Stream	Checklist Items	Before the field	During the field	After the field
	- Screw driver	- Ensure the screwdriver matches the screw type.	- Carry the screw drive and a spare for use during deployment and retrieval	- Store in a locked cabinet.
	- Screw	- Count the number of screws to match the number of sensor casings to be used. 1 per casing.	- Carry appropriate number and spare for use for every deployment.	- Store in plastic back and lock in the cabinet.
	- Rubber fitting	- Count the number of rubber fittings to match the number of sensor casings to be used. 1 per casing.	- Carry appropriate number and spare for use for every deployment.	- Store in plastic back and lock in the cabinet.
	- Strings(Varying in color:Pink,yellow,blue)	- Check the quality and quantity of strings. Identify strings that are different in color. (Choose colors that are fair)	- Carry the threads separately to avoid entanglement. - Attach the threads on the sensors.Each thread color to a different household.	- Remove all the strings from the sensors and dispose of the threads following standard waste management practices.
	- Plastic container(with 3 compartments)	- Make sure you have adequate containers for deployment.	- Carry the appropriate amount of containers for deployment. - Train the guardian on how they should use and store the sensors inside the containers.	- Store the containers in a locked cabinet.

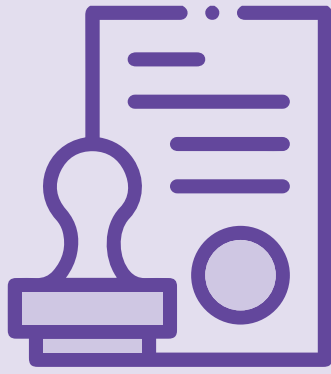
Data Stream	Checklist Items	Before the field	During the field	After the field
	<ul style="list-style-type: none"> <li>- Batteries</li> </ul>	<ul style="list-style-type: none"> <li>- Identify high quality batteries for deployment.</li> <li>- Conduct an in-house test with the batteries by inserting them in 1 sensor and recording the lifespan of the battery.</li> </ul>	<ul style="list-style-type: none"> <li>- Carry appropriate batteries and use them during the deployment.</li> </ul>	<ul style="list-style-type: none"> <li>- Dispose the batteries following the recommended waste management practices.</li> </ul>
Tablet		<ul style="list-style-type: none"> <li>- Install the Jotform</li> </ul>	<ul style="list-style-type: none"> <li>- Charge and use it during every deployment.</li> <li>- Capture the participants metadata and scan the sensor barcode using the Jotform.</li> </ul>	<ul style="list-style-type: none"> <li>- Make sure all Jotforms have been synced.</li> <li>- Store in a locked cabinet.</li> </ul>
Programmer		<ul style="list-style-type: none"> <li>- Check condition</li> <li>- Ensure you have</li> </ul>	<ul style="list-style-type: none"> <li>- Use the Programmer to wake up and timestamp the master sensor</li> </ul>	<ul style="list-style-type: none"> <li>- Use the programmer to download the data from the sensors.</li> <li>- Wrap in bubble wrap and store in a locked cabinet.</li> </ul>
Laptop		<ul style="list-style-type: none"> <li>- Ensure you have the socio patterns web app installed in the laptop.</li> </ul>	<ul style="list-style-type: none"> <li>- Use the laptop to timestamp the master sensor</li> </ul>	<ul style="list-style-type: none"> <li>- Use Pc to download and upload sensor data to the aws bucket.</li> <li>- Store in a locked cabinet.</li> </ul>
Multi-module face-to-face survey		<ul style="list-style-type: none"> <li>- Ensure you have the survey CTO application installed.</li> <li>- Ensure the data collectors are assigned enumerator Ids</li> <li>- Ensure airtime is loaded for sending and downloading data.</li> </ul>	<ul style="list-style-type: none"> <li>- Charge and use it for every session.</li> <li>- Synchronize records after every session.</li> </ul>	<ul style="list-style-type: none"> <li>- Ensure all records are synced to the survey CTO dashboard.</li> <li>- Store in a locked cabinet.</li> </ul>

# 11

---

Appendix A:

## Legal and ethical considerations



**Authors:**

Kerstin Noëlle Vokinger,  
Nicolò Tomaselli, and Guilherme Lichand

## 11. Legal and Ethical Considerations

Data collection involving biomarkers requires even higher ethical standards than normal research with human subjects. In the context of the Child Development Study developed in Malawi between 2020 and 2023, the following legal and ethical framework has been adopted. We report it here so that future researchers and implementers can draw inspiration and consider the importance of having clear rules and guidelines on how to handle privacy, confidentiality, and sensitive data.

It is worth noting that the Child Development Study has also been endorsed by the Malawi Ethical COMREC and the Institutional Review Board (IRB) of the University of Zurich. Any project involving the collection of data on children, especially with wearable devices, must be approved by these types of committees. Informed consent and data protection policies must comply with relevant national and international legislation.

Another good practice is to conduct an accompanying ethnographic study. In the case of the Child Development Study, a study was conducted to confirm that children were not exposed to risks and were mostly happy to participate in data collection with sensors. The study found that families were happy to participate in the wearable technology stream and that parents showed no hesitation towards the use of the technologies (van Waes, forthcoming).

\*\*\*

### **Ethics Framework for Research on Child Development Based on Wearable Technologies**

#### **Preamble**

The ethical framework shall safeguard the individuals' rights and provide guidance for the study "Research on Child Development Based on Wearable Technologies" ("study"), conducted by the economics department of the University of Zurich in collaboration with the Malawi College of Medicine and UNICEF.

All parties express their mutual will to abide by this ethical framework.

#### **I. General Provisions**

##### **1. International and National Laws**

The following European and International Conventions shall be recognized in this framework and the study:

Convention on the Rights of the Child of 2 September 1990

European Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data of 28 January 1981

European Convention for the Protection of Human Rights and Fundamental Freedoms of 4 November 1950

International Covenant on Economic, Social and Cultural Rights of 16 December 1966

International Covenant on Civil and Political Rights of 16 December 1966

International Convention of the Elimination of All Forms of Racial Discrimination of 21 December 1965

Convention on the Elimination of All Forms of Discrimination against Women of 18 December 1977

Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment 26 June 1987

Convention on the Rights of Persons with Disabilities guarantees all human rights for people with disabilities 13 December 2006

Convention for the Protection of All Persons from Enforced Disappearance of 18 December 1992

Framework Convention for the Protection of National Minorities of 21 October 1998

African Union Convention on Cyber Security and Personal Data Protection of 27 June 2014

Additionally, Swiss law and principles shall be considered in the conducted study in order to protect the individuals' rights.

## **2. Purpose and Scope**

The following provisions shall protect the individuals' rights.

## **3. Définitions**

Participants: Persons – children and legal representatives – who participate in the study.

Children: Persons who have not yet reached the age of maturity (18 years).

Legal representative: Person(s) who are legally responsible for a child (usually but not necessarily the caregiver).

Power of judgment: The ability to form (intellectual component) and declare (voluntative component) an own will.

Personal health data: All data of individuals that can directly or indirectly provide information about the health status (physiological or pathological status) of an individual person.

Health data: All data of individuals that can directly or indirectly provide information about the health status (physiological or pathological status) that cannot anymore be linked to the individual person due to the anonymization process.

Wearable Technologies: All devices used in the study that continuously collect health data.

Anonymization: Protecting the confidentiality of study participants through data anonymization, which is a process that reduces the risk of re-identifying individuals within a given dataset. After the anonymization process, the data cannot anymore be linked to the individuals and the data is no longer considered as “personal health data”, but rather “health data”.

## 4. Basic Principles

Non-discrimination: All parties agree that no participant in the study will be discriminated against due to their origin or descent, race, gender, age, language, social standing, chosen way of life, as well as their religious, ideological or political views or physical, mental or intellectual capacity.

Voluntariness: All parties acknowledge that participation in the study or eligibility to be part thereof is purely based on the participant’s voluntary choice. No person can be forced to:

- o participate in the selection process or quit the selection process for the study if they do not choose to do so based on their own free will; or

to participate in the study or quit the study if they do not choose to do so based on their own will after they were selected as participants.

Informed consent: The informed consent-principle ensures that participants are informed about all important steps of the study, agree to be part of the study and are aware of the rights they forfeit upon enrollment in the study.

Transparency: Participants have to be kept informed about the progress of the study as well as the current phase they find themselves in. To that end, study participants must have the opportunity to contact a responsible person of the study. Furthermore, the participants must be informed about the collection and use of their health data upon enrollment in the study as well as throughout the run of the project.

Data processing and prohibition of commercialization: The collected health data is only to be used for research purposes with immediate relation to this particular study. It is especially forbidden to use the collected health data for commercial or monetary purposes.

Anonymization: The risk of re-identification through data linkage is essentially unpredictable because it can never be assessed with certainty what data is already available or what data may be released in the future. It is also generally infeasible to see data return (e.g. the right of individuals of recalling data or removing it from a website) as a safeguard given the difficulty, or impossibility, of securing the deletion or removal of data once it has been anonymized and/or made public.

## II. Information and Consent

### 1. Information and Consent Process

Every participant is being informed in an oral information session about the study, the collection and processing of the personal health data, and the voluntariness of the consent.

Every participant is provided with a written document (consent form: name of the document) about the most relevant information of the study.

Every participant signs the informed consent form [name of the document] in order to express his/her consent to the study participation. In case the participant is illiterate, the participant leaves a thumb stamp.

## 2. Role of Participants and Legal Representatives

It is assumed that participants until the age of 9 do not yet have the power of judgment. For participants under the age of 9, the legal representative has the responsibility for giving the consent. In cases where participants have already reached the power of judgment, both the participant and the legal representative are being informed and giving their consent.

For participants between age 9 and 18 both, the participant and the legal representative are being informed and give their consent.

## 3. Right to Refuse

Every participant may refuse to participate in the study once they have been selected to be part of it.

Every participant has the possibility to opt out of the study anytime.

## 4. Confidentiality

The participation or opting out of the study is confidential and for internal use only. The identity of the participants and related are not to be shared with anyone except the project personnel.

## III. Research with Wearable Technologies and Data

### 1. Wearable Technologies

All applied wearables in the study collect health data of the individuals.

All applied wearables in the study are of high quality.

All applied wearables in the study do not lead to any medical harm of the individuals.<sup>1</sup>

### 2. Data Collection, Processing, and Use

The scope of health data collection is limited to the data needed for the study.

All personal health data is treated confidential.

Technical measures are being implemented to anonymize personal health data subsequently after the collection.

No personal health data is stored in the datasets that are used for the research. To protect human subjects, de-identification should occur as early as possible in the research process.

The following technical precautions are in place in order to protect the collected data and to avoid access by unauthorized personnel:

- Data is collected in electronic form and subsequently is archived in a cloud-based server.<sup>2</sup>
- All data transfers are end-to-end encrypted thanks to the use of encryption methods.
- Data will be protected while in-transit (as it travels to and from the cloud server) and at-rest

---

1 In case the study will use technological parts (for example, small electronic components) it may occur that children could risk swallowing some of the parts and be exposed to the risk of choking. The research team will make sure to protect small parts and include a warning message in the informed consent..

2 If feasible, the research team will avoid third parties services from the US.



(while it is stored on disks in the data centers).

- All interactions in the dashboard and in mobile applications use HTTPS encryption.
- Only authorized users with token actions are available for mobile application and dashboard.
- All devices used in the field for data collection and all computers used by researchers in the context of the Child Development Study are password protected and avoid any local storage of any information (e.g. information is stored on cloud-based servers).

Participants (and their legal representative) may have access to their personal health data.

Participants (and their legal representative) do not have access to anonymized health data.

Personal health data and anonymized health data may only be used for this study; secondary use of the data is prohibited.

Both, personal health data and anonymized health data may not be shared with third parties.

Commercialization of personal health data and anonymized health data is prohibited.

#### **IV. Additional Requirements for Third Parties**

Manufacturers of the wearable technologies have the right to get access to aggregated, anonymized health data and the study results.

Manufacturers of wearable technologies do not have the right to get access to personal health data.

Governments have the right to get access to aggregated, anonymized health data and the study results.

Governments do not have the right to get access to personal health data.

#### **V. Supervision**

Malawi College of Medicine and the Center for Child Well-being and Development at the University of Zurich are supervising the study. The study will obtain (and renew when necessary) IRB approvals from competent bodies in Malawi and Switzerland.

# 12

---

Appendix B:

## Vendor TOR template

# 12. Vendor TOR template

[TEMPLATE]

Agreement for Services Provided Under a Contract for Work and Services

between

[...]

as Purchaser

and

[...]

as Service Provider

concerning

The development of devices and software solutions to measure biomarkers in children from one to [...] years old

## 1. Object of the Contract and Description of Goods/Services

The Service Provider undertakes to successfully complete the following contractual services against payment:

The Service Provider shall develop a device for biosignal measurement pad device, which shall be able to measure children from x to y years old. The device will be able to measure ECG signals. The device shall have expanded battery life (e.g. expected battery life is estimated in at least 300 minutes of continuous utilization -- the device will have a port to plug external batteries like power banks) and simple charging via micro-USB. The Service Provider will manufacture 10 samples of this device.

The Service Provider shall develop updates to existing software infrastructure that will allow using it together with the device. These updates will include updates for the mobile application for an Android tablet, which will allow the Purchaser to automatically retrieve, record and synchronize electrocardiogram (ECG) data and signals. These updates will also include updates for the cloud infrastructure that will allow the Purchaser to process the collected ECG data. The development of the device and manufacturing of these [...] samples of it will occur during the device development stage.

Apart from the development of the device and manufacturing [...] samples of it (development stage), this part of the project would also include possible improvements of the devices that would take up to 200 hours of time for team members (modification stage). In case of substantial modifications (more than 200 hours of work of team members), they would be subject to another billing, according to SLA terms from this document. In case of light modifications (less than 200 hours of work of team members) unused time of team members can be used for realization of other updates, according to the written (email) requests by the Purchaser team members and scope of the project defined in this document.

Delivery kit for each device includes the device itself and the charging cable for it (3rd-party charging cables will not be supported). External electrodes will be manufactured separately for free for the [...] device samples. Cost of this contract includes shipping and customs clearance costs of the devices.

Additionally, the Service Provider shall develop updates for existing software infrastructure to implement questionnaires created on the CAPI platform into developed software. The structure

of these questionnaires will be partly predetermined. Questionnaires will be created in the CAPI platform and added to the tablet application when there is a stable internet connection. These questionnaires can be used to collect data during measurement sessions according to the current design of questionnaires section of the tablet application. Numerous questionnaires can be available for filling in the application at the same time (all questionnaires for a specific CAPI platform account). Answers to the questions in the questionnaire will be stored in the memory of the tablet until being synchronized with the CAPI platform using a stable internet connection, manually (by pressing a button in the application, on demand).

The questionnaires should have the following predetermined technical elements to be compatible with the application:

- The first question of each questionnaire should be the technical question Enter child Field ID with the answer being a text field.
- The Enter child Field ID field will be filled in automatically by the tablet application. The flow here is the following:
  - Field agent selects a child
  - Field agent selects "questionnaires" option for this child
  - Child identification information is being entered automatically, other questions are being entered manually by the field agent.
- Other predefined questions may be added later by the Team during development of this update
- Not predefined questions of not fixed structure that collect information needed for the researchers can be added to the predefined technical questions to collect needed information (our limitations ensure the work of the update, but the customer can still add all needed questions)

This update will work with the versions of CAPI platform software versions that are proposed at the date of signing the contract. The Team has no responsibility for problems in the performance of this update that may happen due to updates or problems from the CAPI platform side. This update will be delivered according to the limitations of the CAPI platform. If any functionality is not supported by CAPI platform - it cannot be implemented. An example of it can be if the real behavior of the CAPI platform service doesn't match with the one described in the documentation. In such a case, a discussion on how to deal with this problem will be initiated between our parties. This may affect the cost of this sub-project. Filled-in questionnaires will be presented on the CAPI platform and added to the internal cloud infrastructure of the project (web dashboard).

The Service Provider shall also develop updates to the EEG data collection pipeline. This update will include basic functionality of showing the user triggers and recording time when they were shown in the application. During EEG recording, a person is shown pictures (triggers). Pictures to be shown are selected randomly from a set of two pictures, which is being provided once and in advance (before the start of implementation of a software update). Random pictures (picture 1 or picture 2) are being shown at random timestamps:

- The first timestamp happens after a random amount of time has passed after the start of recording. Next, timestamps happen after a random amount of time has passed after a previous timestamp. These random amounts of time are uniformly distributed in a window [1.4 seconds .. 3 seconds].

- At each timestamp, one of the two pictures is randomly selected to be shown. The probability of showing picture 1 is 0.8, the probability of showing picture 2 is 0.2

Pictures are being shown for 0.1 seconds. After this time window, a picture disappears, and no picture is being shown until the next timestamp. Recording continues until picture 2 has been shown 50 times, and in total, 250 pictures have been shown. Data is being stored in a specific format. Each point of digitized EEG data is being recorded as a vector that consists of the following values:

- Timestamp when this point was recorded. The “zero time moment” for the timestamps is selected according to the preferences of a developer.
- All the different EEG values corresponding to different channels
- 0 if at this moment no picture is being shown, 1 if at this moment picture 1 is being shown, 2 if at this moment picture 2 is being shown
- The data format will be identical to an example file being sent to us in advance, before the start of the project.

After the update was developed and tested, additional assistance will be provided to:

- Test if the approach works in practice by:
  - conducting 5 - 40 EEG measurements using the developed software update to collect the data
  - providing data to researchers
  - recruiting up to [...] adult subjects to conduct these measurements
- Updating the app according to the results of testing & limited to modifying parameters and changing the pictures being shown, without changes in infrastructure
- Supporting the processes of analyzing collected data and deployment of the solution for the Malawi-based team
- Additional support via teleconference calls
- All the additional assistance described here is limited to 20 hours of additional time spent by members of the R&D team of the Service Provider.

Additional functionality of adding ECG timestamps to the data would also be developed as part of this update together with the possibility of recording auxiliary Muse channel in the existing infrastructure, which would include software improvements and manufacturing of [...] auxiliary electrodes for Muse according to the guidelines previously received from the EEG analysis team introduced by the purchaser.

Dataset downloading functionality improvements will be developed by the Service Provider as a part of this agreement. These improvements will be focused on providing the possibility to download data without directly using Amazon websites. The user will be able to select the records to download using a web dashboard, then the download link for these records will be generated. This link will be pasted in a specially developed python script and then the data download will start.

IDs management and anonymization functionality will be developed by the Service Provider as a part of this agreement. Existing ID based approach will be replaced by an approach that is based on 3 IDs:

- The internal ID used by the system itself (backend, DB), not shown to users
- Field ID, used in the field, not shown to the researchers in labs. This ID will be generated by the tablet application when the field agent adds a new child to the system (on one of the tablets). Then, when the data is synchronized and uploaded to DB, internal ID and Lab ID for this child (Field ID) are created. After this, during the following synchronization on every tablet, information about this child (Field ID) is being downloaded from the cloud and it becomes possible to use it on other tablets.
- Lab ID, used by researchers in Lab, can't be explicitly linked by them to Internal ID or Field ID to guarantee anonymization
- Links between Lab IDs, internal IDs and Field IDs will be used by the system itself and seen by technical specialists that have access to the system (not by researchers or field agents). System superadmin will be able to see the links via admin dashboard.

CAPI questionnaires will be used to connect Proximity Sensor IDs with the Field IDs of children. The following algorithm will be used:

- Field Agent Selects the child (participant of the experiment). Then the Field Agent is able fill in the Proximity Sensor questionnaires for this particular child.
- Questionnaire 1 - Which proximity sensor was given to this child
- Questionnaire 2 - Which sensor was taken from this child
- Answers - same for both questionnaires - Field ID (automatically), date (automatically), Proximity Sensor IDs (manual input) - both the Hardware ID and the Deployment ID
- This questionnaire is being filled in after the field agent has selected the child from the list of all children. And so, the Field ID and date are known to the tablet application and are being filled in automatically.

Functionality of uploading Proximity Sensor IDs CAPI platform Questionary data on links between Proximity Sensor IDs and Lab IDs into the dashboard will be added. It will be based on using CSV files of a predefined structure: first column - proximity sensor ID, second column - date, third column - Lab ID, Fourth column - was the Proximity Sensor Given or Taken.

The functionality of adding a phone number to the child registration form will be developed by the service provider, as well as the functionality of changing this phone number via the tablet application. For anonymization and data protection reasons, this phone number won't be shown to the researcher in the lab. Only system superadmin will be able to see these phone numbers, match them with Field and Lab IDs and assist the researchers with the phone calls when needed. Superadmin will involve other authorized personnel to manage the handling of PII's when needed. Only Lab ID, Age, Gender, Community and Collected Data information will be shown in the dashboard.

Service provider would guarantee correct work of the android tablet software application only for the Android tablets from a list of models approved by the service provider and only in case of the correct work of particular android tablet.

Purchaser has an intent of launching future projects with the Service Provider, which would include:

- Software update on HRV features calculation & R-peaks detection
- Software update on processing of HRV artifacts
- Software update on ECG Signal Filtering
- Software update on ECG Signal Validation
- Software update on PPG Signal Filtering
- Software update on PPG Signal Validation
- Software update on improved data visualization
- Software update on advanced data visualization and analysis in web dashboard
- Software update on Android & Backend Data Encryption
- Manufacturing of 200+ devices

Possible requirements and costs of these future projects are described in the [...] for this document.

## 2. Elements of the Agreement

[...]

## 3. Instruction

Under section [...], the Service Provider is to provide the following instruction services:

- The Service Provider shall provide documentation in English to describe how the device will need to be handled by the person in charge of collecting data in the field and its maintenance instructions;
- The Service Provider shall provide written standard protocols in English on how to use the device, so that any agent will be able to instruct patients while using the devices;
- The Service Provider shall provide written documentation in English on how to use the dashboard for the integration of different signals coming from different devices;
- The Service Provider shall make sure that ad-hoc technical services in English are available any time the Purchaser has difficulties in using the device or generating data from the device. Ad-hoc technical support shall be available during the - and this during the agreement and for three months after the test and acceptance of the deliverables. These follow-up services shall be provided via the use of teleconferences or other long distance communication devices.
- Ad-hoc technical support that exceeds 200 hours of the time of team members of the Service Provider (as defined by the description of modification stage of segment 1 of this document) may be billed according to the SLA (annex 1) at the discretion of the Service Provider in case of the Service Provider representatives notifying Purchaser representatives of such billing in advance, before providing particular ad-hoc technical support activities subject to such billing.

## 4. Deadlines

[...]

## 5. Payment

[...]

## 6. Duties on the Part of the Purchaser

The Purchaser has the following duties to cooperate (this list is exhaustive).

- To participate in project progress meetings;
- To grant physical access rights and system authorizations, when necessary;
- To conduct testing and acceptance procedures following implementation;

Only if the obligations have not been fulfilled, and in cases of force majeure, may the Service Provider be exempted from rendering the performance or part-performance concerned. However, this exemption applies only if the Service Provider is not at fault in any way for the failure to fulfill the obligation, and to only that extent to which performance is rendered impossible or excessively difficult. Where this is the case, the Service Provider must notify the Purchaser immediately. In such cases, the Service Provider is obliged:

- to render all of those elements of performance which are unaffected by the failure to fulfill the obligation;
- to do all that it can to render that part of performance that is affected at least to some extent; and
- as soon as possible to resume the performance that is affected.

The deadlines that have been set will be deferred in accordance with the length of time for which the obligations are not fulfilled or the force majeure persists.

## 7. Testing and Acceptance

Joint testing is carried out prior to acceptance. A protocol is kept of the tests and their results, and is signed by both Contracting Parties.

The criteria of acceptance are the following:

- Devices (hardware) and System (Software) accomplished satisfactory their functions and application for the demands stated in the present Agreement (before the pilot phases);
- Devices (hardware) and System (Software) substantially show good performance, user experience and robustness during the pilot, with the extent of bugs, system crash or glitches that won't affect the capacity to collect data in a continuous way (e.g. max 10% of downtime during the whole experiment);
- Data reliability on collecting, storing, retrieving.

Functions of the devices that will be tested:

1. Signals: Electrocardiogram.
2. Data transmission: the capacity to transmit the data packages using Bluetooth connectivity.
3. Data quality/format: All the data from the cloud system will be downloaded using a CSV format, with the proper structure (columns and rows) and the researcher in charge should be informed about any modification of that structure.

Definitions of defects: The Purchaser classifies defects as "significant" or "insignificant". Defects classified as "insignificant" shall not prevent acceptance. If significant defects are found, acceptance is postponed. The Service Provider rectifies the defects found immediately and invites the Purchaser to take part in further testing in good time.





## CONTACT

Kamuzu University of Health Sciences <https://www.kuhes.ac.mw/>

Center for Child Well-being and Development. [www.ccwd.uzh.ch/](http://www.ccwd.uzh.ch/)

University of Zurich, Department of Economics. [www.econ.uzh.ch/en.html](http://www.econ.uzh.ch/en.html)



**KAMUZU  
UNIVERSITY  
OF HEALTH SCIENCES**



University of  
Zurich <sup>uzh</sup>

Department of Economics

**Stanford**

Center on Early Childhood



**CCWD**  
CENTER FOR CHILD WELL-BEING  
& DEVELOPMENT