CHAMACOS Data Sharing Policies

This document outlines the policies for sharing and use of data and samples generated by the Center for Health Assessment of Mothers and Children of Salinas (CHAMACOS) study at the UC Berkeley School of Public Health Center for Environmental Research and Children's Health (CERCH). With these policies, we aim to enhance collaborations with UC Berkeley faculty and students and outside investigators while avoiding misunderstandings and conducting high quality science as efficiently as possible.

I. Availability of CHAMACOS Data

The CHAMACOS Dataset comprises a rich trove of data and samples spanning over 20 years of chronological data collection (1999-2020), or up to 19 years (first trimester gestation – age 18y) of development for individual participants. The CHAMACOS data portal [https://cerch.berkeley.edu/investigators] offers the most complete and up-to-date summary of data and samples available in this dataset, as well as a general overview of how and when participants were recruited and demographic characteristics of maternal and youth participants. Researchers interested in accessing CHAMACOS data are also encouraged to review our many CHAMACOS publications to learn more details of how data of interest were collected.

It should be noted that the quantity of data and samples available for sharing (i.e., the number of participants on whom we can provide data) may be less than the number included in our own publications. One important factor that will influence shareable sample size includes the level of consent provided by individual participants. With regards to biosamples and environmental samples, the number of samples available to share will be affected by how much of each sample has been consumed by previous analyses. Finally, we want to highlight the distinction between the "CHAM1" and "CHAM2" portions of our cohort. CHAM1 participants were recruited in the prenatal period. CHAM2 participants were recruited at age 9y. Prenatal and early childhood biological and environmental samples and early health and development data are only available for CHAM1 participants.

II. Requesting CHAMACOS Data

Anyone wishing to access CHAMACOS data or samples, or to enrich the existing dataset through new data collection, must first submit an Initial Data Use Application (see **Appendix 1**). Initial Data Use Applications can be submitted at any time, and should be submitted simultaneously to:

- Katherine Kogut, CHAMACOS Study Coordinator, at kkogut8@berkeley.edu and
- Kyna Long, CHAMACOS Data Manager, at kzlong@berkeley.edu

Please include "Initial Data Use Application" in the subject line.

III. Decision-Making Authority and Approval Timeline

Initial Data Use Applications will be reviewed on a monthly basis by the CHAMACOS Executive Committee (**Table 1**), which has primary decision-making authority. In some instances, when a core member of the Executive Committee is not available for an extended period of time and a decision must be made in their absence, alternate Investigators may be asked to serve in their place.

Table 1. CHAMACOS Executive Committee				
Core Members:	Brenda Eskenazi (PI)	Asa Bradman	Nina Holland	Kim Harley

Possible Alternates:	Kim Harley or Asa Bradman (with alternates their roles)	Robert Gunier	Karen Huen Andres Cardenas	Sharon Sagiv, Ana María Mora, Julie Deardorff

When reviewing Applications, the Executive Committee will weigh the following factors:

- Scientific merit;
- Achievability (including turnaround time required by the Applicant);
- Overlap with funded grant aims;
- Human subjects considerations, including consistency with consent forms signed by participants, request for de-identified versus potentially identifiable data elements;
- The **cost** of providing the data relative to available funding. (See section IV on Costs); and
- (For biosample request): the availability and uniqueness of a biosample relative to the scientific merit of the proposal

The Executive Committee may decide to either: 1) accept the proposal (as is or provisional upon reimbursement for costs), 2) provide feedback and invite the investigator to revise and resubmit, 3) request input from more investigators, or 4) reject the proposal.

In some cases, the input of other investigators will be a necessary prerequisite to acceptance of a proposal. This is the case when the proposal will make use of a subset of data that has its own Data "Owner" (i.e., the PI of the grant which funded the data collection), or when it pertains to a subject area for which we have a local subject matter expert. The justification for including local experts in the decision-making process is two-fold: 1) they are in the best position to assess the strength and feasibility of a proposal, and 2) in some cases, we will want to insure that certain concepts are analyzed in a specific way, particularly if the project will result in publication. (For example, we want concepts such as "early adversity" in CHAMACOS to be presented with consistency across publications). In cases where the input of other investigators is necessary, the Executive Committee will determine how best to obtain their input. In simple cases, emailing the Initial Data Use Application to the other investigators and receiving their feedback via email may be sufficient. In more complex cases, it may necessary to postpone review of the Initial Data Use Application until a second meeting at which the other investigators are present.

Tables 2 and 3 provide current lists of Data "Owners" and Subject Matter Experts, respectively. In most cases, the "Data Owner" is the PI of the grant that allowed for the collection of those data. (Note: in some cases data may have 2 owners, for example, if one grant collected a biosample and another analyzed it for certain chemicals.). The "Subject Matter Experts" are the co-Investigators who were integrally involved in the data collection and are experts in that topic. These lists are neither exhaustive nor permanent and may be added to as new data are collected. Data "ownership" status may change after a certain period of time, such as 5 years after the end of the grant in question or the active use of the dataset.

In most cases, we envision a consensus among the Executive Committee, and any additional Data "Owners" and Subject Matter Experts. However, should an impasse exist, Brenda Eskenazi maintains final veto power until at least 12/31/2025. Over and above this, if the grant which funded the collection of requested data or samples is still active, the Data "Owner" (i.e. the PI of the active grant) also maintains veto power.

Table 2. Data "Owners"

Specialty Data Type	Data "Owner"
Child and maternal data (pregnancy to age 12yr)	Brenda Eskenazi, with Kim Harley as a Possible Alternate
Biomarkers of effect and susceptibility from maternal pregnancy, and children from birth to 14 yo, including epigenetic, immunological, metabolomic, microbiome and functional genomics markers	Nina Holland, with Karen Huen as a Possible Alternate
14yr data, especially related to the Trier Social Stress Test (TSST), saliva/hormones, family risk and protective factors, youth risk behaviors, and Tanner staging	Julianna Deardorff, with Megan Johnson as a Possible Alternate
14yr data, especially related to diet/physical activity, actigraphy, and blood samples/results	Nina Holland, with Karen Huen as a Possible Alternate
14yr Social Responsiveness Scale data	Sharon Sagiv
16yr data related to family risk and protective factors, youth substance use and sexual risk-taking behaviors	Julianna Deardorff & Brenda Eskenazi, with Megan Johnson as a Possible Alternate
16yr data related to youth delinquent and criminal behavior, driving, and academic performance	Brenda Eskenazi
fNIRS data (CHAM2 16yr "Pilot" Data and 18y full-cohort data)	Brenda Eskenazi, with Sharon Sagiv and Ana María Mora as a Possible Alternates
18yr data	Brenda Eskenazi, with Ana María Mora as a Possible Alternate
CHAMACOS Youth Council (CYC) Data, including from HERMOSA or COSECHA or LUCIR	Kim Harley
Exposure or Sub-Study Data, including from CHAMACOS Validation Study (CVS), intervention studies, or Organic Diet Study	Asa Bradman, with Rosemary Castorina or Robert Gunier as a Possible Alternates
Phthalates	Nina Holland
BPA and phenols	Kim Harley

PBDEs	Brenda Eskenazi, with Asa Bradman as an alternative
OP Flame Retardants	Asa Bradman
Pesticides	Asa Bradman, with Brenda Eskenazi as an alternative

In most cases, Applicants can expect a final decision from the Executive Committee **within 8 weeks** of submitting their Initial Data Use Application.

Table 3. Subject Area Experts	**Expert**
Autism/Pervasive Developmental Disorder	Sharon Sagiv
Cortisol and other Salivary Hormones	Elizabeth (Birdie) Shirtcliff, with Megan Johnson as a Possible Alternate
Early adversity	Julianna Deardorff, with Megan Johnson as a Possible Alternate
fNIRS	Sharon Sagiv, with Ana María Mora as a Possible Alternate. Also Allan Reiss, Joseph Baker, and Jennifer Bruno from Stanford CIBSR
Genetics/Epigenetics/Biomarkers/Omics	Nina Holland, with Karen Huen or Andres Cardenas as Possible Alternates
GIS mapping	Robert Gunier
Puberty	Kim Harley, Julianna Deardorff
Autonomic Nervous System (ANS) data collected prior to age 7yr	Abbey Alkon
Neurodevelopment	Brenda Eskenazi, with Ana María Mora and Sharon Sagiv as Possible Alternates
Growth and obesity	Kim Harley, Marcy Warner, Nina Holland
Respiratory outcomes	John Balmes, with Kim Harley as a Possible Alternate
Birth outcomes	Brenda Eskenazi, Kim Harley
Reproductive outcomes	Kim Harley, Brenda Eskenazi

IV. Costs to Access CHAMACOS Data and Biosamples

Providing data and samples requires staff time (e.g. to compile datasets, pull samples, seek IRB approval to release data, prepare data/sample transfer agreements) and may also incur additional expenses (e.g. to ship samples). In many cases, it will only be feasible to provide data or samples if we will be reimbursed for doing so. In all cases, the "cost" of the request relative to available institutional resources will be considered as a factor in review of Initial Data Request.

We expect that these costs would be covered by a grant written specifically for the purpose of analyzing these data or samples. Very few exceptions to these costs will be made. Costs may be waived for current CHAMACOS investigators or students on a case-by-case basis.

A. CHAMACOS Data

With regards to all requests for CHAMACOS data, the following costs will be considered in addition to any applicable biological sample use charges outlined above:

Category 1: Secondary analysis of existing CHAMACOS data

The work involved for CHAMACOS staff would include:

a) Obtaining relevant IRB approvals;
b) Preparing new datasets, including new variable creation and/or scoring as needed;
c) Providing context on how data were collected, drafting methods sections, etc;
d) Assisting in preparation of project progress reports and publications.

The cost for this category of work would be calculated as at least 5% of Study Coordinator's time, and at least 10% of Data Manager's time, possibly decreasing to 5% time in subsequent years after dataset has been created. In addition to the actual staff time, the grant should cover the proportion of office maintenance costs (e.g. Phone, photocopy, office supplies, software licenses, computers) attributed to those staff members, which we have calculated as $2000/staff member/year. In other words, if a grant covers 10% of the Data Manager's salary + 5% of the Study Coordinator's time in year 1, an additional $300 ($200 for DM + $100 for SC) should be added to cover office maintenance costs for those staff people. Final costs should be determined in consultation with CERCH staff and may vary by project

Category 2: New data collection from CHAMACOS participants

The work involved for CHAMACOS staff would include:

a) Assisting with grant writing;
b) Obtaining IRB approval;
c) Assisting in preparation of data collection protocols and instruments;
d) Coordination of data collection staff;
e) Facilitating contact with participants;
f) Management of incoming data;
g) Preparation of new datasets;
h) Merging of new and existing data;
i) Assisting in preparation of project progress reports and publications.

The cost for this category of work would be calculated on a project-specific basis, but is likely to include a significant portion of the following people's time: Study Coordinator, Data Manager, Data Analyst, Administrator. A portion of the Key Investigator time will also be essential. Please refer to Tables 2 & 3 for appropriate investigators to include. Members of the CHAMACOS Executive Committee would need to be involved in crafting the project budget prior to submission to the funding agency.

B. CHAMACOS Samples in the SPH Biorepository

With regards to requests for biological samples, costs will be considered as follows:

Category 1: Use of Existing Biosamples.
The work involved for CHAMACOS staff would include:

a) Assessing sample availability based on type, aliquot numbers, volumes;
b) Preparing summaries on sample availability that require extensive queries; and
c) Pulling and preparing select samples for shipping/analysis.

The cost for this category of work would likely be calculated as 10% of Dr. Karen Huen's time for 1 month or longer depending on the complexity of the request, plus 25% of a GSR's time for 1 month (assuming ~100 samples) or longer if many more samples/types of specimens are included.

Note: in some cases, we might have to also charge for complicated queries about the available data or for pilot analyses if the work involved for CHAMACOS staff is excessive.

Category 2: Supporting New Sample Collection and Processing (i.e. biorepository services)
The work involved for CHAMACOS staff would include:

a) Preparing new sample collection/processing protocols, and
b) Receiving, processing, and storing new samples according to these protocols.

The cost for this category of work would be calculated as 20% of Dr. Huen's time during the first month of the grant to establish protocols and 10% per month thereafter for sample processing and storage. We will also charge 25-50% of a 1 month minimum GSR.

Category 3: Sample Analysis to be Conducted by Holland Lab (i.e. research activities)

This would include in-house biomarker analysis to be conducted directly by Dr. Holland's lab, under the direction of Dr. Huen.

The cost for this category of work would be calculated as at least 20% of Dr. Huen's time depending on complexity of analyses and at least 50% GSR time for one month or longer. It would also include any materials and testing supplies required for the analyses.

Charges can be paid via a subcontract to CERCH and/or Dr. Holland's Lab through the new grant funding the research, by contract or for small amounts by "fee for service".

V. Timeline and Steps to Receive CHAMACOS Data Once Approved

Once an Initial Data Request Application is approved by the Executive Committee, **the timeline to receive data can range from as little as 1 week to as long as 3 months (or longer if outside funding or IRB approval is required).** The timeline for receiving data is dependent on the following factors:

- Whether the applicant is at UC Berkeley or is from an external institution, which impacts the complexity of IRB approval and the need for a Data Transfer Agreement;
- The degree of IRB approval needed, from what IRB(s), and the timelines of those IRBs;
- Whether funding is to be provided, and the timeline in which it is provided;
- Whether biosamples are requested;
- The complexity of the data request, including the number of variables requested and how well established these variables are. (Variables which we have already published upon will generally be "clean" and ready to share, whereas new variables may require a good deal of time to clean, score, create, etc.); and
- Competing demands on staff time.

Upon approval, the Applicant will be asked to submit the following to the CHAMACOS Data Manager:

- A Data Request Form detailing the specific variables they are requesting (see **Appendix 2**),
- A Confidentiality Agreement (**Appendix 3**), and
- An Initial Data Use Agreement (**Appendix 4**) in which they acknowledge conditions of use of CHAMACOS data as detailed below. (They may later be required to sign a more formal Data Transfer Agreement, if applicable).
- The Applicant may also be asked to provide proof of completion of CITI training (regarding protection of human subjects).
- A timeline for completion to bring to publication

The CHAMACOS Data Manager and Study Coordinator will advise Applicants by email of what is needed from them, as well as a reasonable estimate of how soon data will be provided.

VI. Special Conditions for Use of CHAMACOS Data

A. CHAMACOS Investigators and Staff Must Provide Oversight of Work

CHAMACOS Investigators are committed to ensuring that all publications or professional presentations based on CHAMACOS data are of the highest quality. We wish to ensure that the methods accurately reflect how data were collected in the field and that the data analysis and conclusions are appropriate. The level of involvement of CHAMACOS investigators will be dependent on several factors, including:

- The experience and training of the researcher using the CHAMACOS data,
- Whether the analysis is being conducted for the researchers' own academic purposes only (e.g., class project, thesis, dissertation) or is intended for publication or presentation to a professional audience
- Whether CHAMACOS data are the main focus of the analysis or represent a relatively small portion of a pooled or meta-analysis, and
- The "age" of the variables being used (i.e., we may seek more active involvement when variables are newer and have not been used extensively in previous analyses.)

Possible oversight by CHAMACOS investigators may include:

- Sharing unique knowledge of how the data were collected and special circumstances in the field.
- Selecting which variables to provide researchers (e.g., to ensure consistency with previous publications, limit sharing of identifiable, or sensitive data);
- Advising on covariates and confounding. (Usually this will require researchers to submit a Directed Acyclic Graph (DAG) illustrating the hypothesized interrelationships of dependent and independent variables and covariates.)
- Requiring researchers to present preliminary findings to CHAMACOS investigators at a Works in Progress (WIPs) meeting (which can be attended in person or remotely by the researcher)
- Requiring researchers to present "final" presentations to CHAMACOS investigators at a WIPs meeting prior to presentation to professional audiences
- Requiring code review by another qualified analyst

B. Authorship

In nearly all cases CHAMACOS investigators will serve as coauthors on publications, and in this role will review manuscripts prior to journal submission. Authorship should be discussed and determined early in the process. Data "Owners" and Subject Matter Experts

should be offered authorship. Other possible authors include CHAMACOS Data Managers/Analysts, Study Coordinators, and Field Staff. We will follow the journal's guidelines for authorship requirements. Any questions about authorship should be brought to the Executive Committee.

C. Code Review

Before papers are submitted for publication, the data analysis code that creates all the tables and figures in the paper must be submitted to the Data Manager for code review. The lead author must submit a suggested person who is qualified to review the code, ideally a co-author.

D. Commitment to Community

CHAMACOS is a community-based participatory research project and community input is a critical facet of our work. We have an active Community Advisory Board (CAB) that includes participants and representatives of the farmworker and grower communities. We are committed to providing CAB members the opportunity to review our work prior to publication. All researchers who intend to publish using CHAMACOS data must share the abstract, and, if requested, the complete manuscript, at the time of initial submission to a journal for publication. Embargo is respected by the CAB. The abstract should be submitted to Dr. Asa Bradman: abradman@berkeley.edu..

E. Paper Submission

Before a paper can be submitted, the lead author (including internal CHAMACOS investigators) must bring it to the attention of the Executive Committee. The Executive Committee will:

- Ensure that the CAB reviews the paper
- Decide whether a press release is indicated
- Ensure that authorship is appropriate
- Ensure that acknowledgements, including grant funding, are correct
- Review the costs of publication and ensure that there is a source of funding if the journal requires page charges.

F. Press Releases

All press releases concerning CHAMACOS findings should be coordinated with the UC Berkeley press office. If the lead author is from an outside institution, that institution's press office must be in communication with UC Berkeley's Office of Media Relations. Similarly, if the journal chooses to have a press release, that must also be coordinated with UC Berkeley's Office of Media Relations. Both the Executive Committee and the CAB need to be informed of the press release and be given the opportunity to review the press release at least 2 weeks before its publication. Dr. Asa Bradman (abradman@berkeley.edu) will coordinate communication with the CAB.

G. Mechanisms of Data Transfer

Data will be transferred from the Data Manager via a secure file on Box.

H. Data Security

Data are not transferable to other analysts or colleagues without written approval. Data are for use on the approved research question only. Data cannot be used by the same analyst for additional analyses without an approved new Data Application.

I. Submission of Final Data Analysis Files to CHAMACOS

All researchers who publish CHAMACOS Data are required to provide a final pdf of the publication, final dataset, and final analysis files (clearly organized by tables and figures in the paper) to the CHAMACOS Data Manager within six months of publication. If the investigator has created new variables, they need to be shared with the Data Manager and a data dictionary provided. These new variables will become part of the CHAMACOS dataset and may be used by other investigators in the future. If the analysis is not published (e.g. student capstone/thesis/dissertation), the final dataset and analysis files (clearly organized by tables and figures) must be provided to the Data Manager within six months of graduation.

J. Deletion of CHAMACOS Data

At the time that the datasets are submitted to the Data Manager, all researchers are also required to delete CHAMACOS data from their personal computers. This must occur within six months of publication (or within six months of the end of the capstone/thesis/dissertation project, for students who do not intend to publish).

K. Protracted Timeline and Extensions

A timeline for publication is required as part of the Data Application Form. If the paper is not submitted by the required time, the researcher may request a 6 month extension. The Executive Committee will determine whether the extension will be granted. In the case of repeated extensions or if the investigator is unable to complete the paper to publication, we reserve the right to take back the data and allow other students or researchers to complete and publish this work. The initial researcher will be required to provide the latest versions of analysis files and data. The initial researcher may not remain as first author on the manuscript if another author must bring the paper to completion.