**NIH R01 Data Management and Sharing Plan 9/12/23**

**1. Data types.** The proposed research for public-use data sharing will include three data types: survey data, stress biomarker data, and activity-space data. The **final** survey dataset will include *self-reported demographic, behavioral, social* (e.g. exposures to adversity, including homelessness, violence, incarceration, social network, etc.) and *health-related data* (e.g. psychosocial measures, mental and physical health symptoms/disorders and diagnoses, health services). The **final biomarker dataset** will include data on the *cortisol concentration* (hormone) collected from hair, *inflammatory proteins* (high-sensitivity C-reactive protein and pro-inflammatory cytokines) collected from venous blood, and *urine drug toxicology screen* (qualitative results for the presence of specific illicit drugs). The **final activity-space dataset** will include data on the characteristics of the places and locations where youth spend their time, including *objective data* (e.g. structural and compositional measures, such as proportion of residents in poverty, residential instability, vacant housing, etc., drawn from the American Community Survey, the Centers for Disease Control and Prevention Social Vulnerability Index, etc.) and *subjective data* (youth self-report on characteristics of the routine locations where they spend time, such as perceived safety, social/physical disorder, social cohesion, and discrimination).

The data will be deposited with the National Addiction & HIV Data Archive Program (NAHDAP), which is supported by the National Institute on Drug Abuse as a data repository. The available data provided from the 300 study participants across the six time-points of collection will be made available in preferred formats (e.g., CSV, SAS, SPSS) and securely transferred to the NAHDAP. Identifying and personal health information will NOT be shared to protect participant confidentiality, including data related to birthdates, names and GPS coordinates of youth routine locations (only characteristics of these locations will be made available), and names and contact information of network members. Participants will be asked to provide informed written consent (18-24 years) or assent (14-17 years) for data sharing. If a minor participant turns 18 years old during the study, we will review the adult consent form with them, and if they consent to continued participation, we will have them sign the form. However, prior to requesting informed consent/assent from participants, the informed consent/assent forms will be reviewed and approved by the OSU IRB to ensure data may be shared as proposed. Protocols and details regarding the data collection, assay techniques, data transformation (e.g. composites created, log or Z-score transformations) and analysis will be included and available through plain text README documents.

**2. Tools, software and/or code.** Data collection will be done in REDCap and data analysis for the project will be performed using a variety of software programs, including SAS, SPSS, STATA and MPLUS. Program software from Colectica Designer will be rented for the 3 members of the data curation team to facilitate documentation of data collection, survey specification, dataset descriptions, and creation of codebooks. Data will be made available in ICPSR preferred formats (e.g., CSV, SAS, SPSS) and securely transferred to the NAHDAP for public access. As noted, all publicly available data accessible through the National Addiction & HIV Data Archive Program will be de­identified and no PHI data will be included. Questions related to the data collection and/or analysis can be sent via email to the principal investigators.

**3. Data standards**

A data dictionary will be provided for the datasets that defines column headers, units of measurement and other pertinent metadata as necessary to understand and reuse the datasets. Submitted data will conform with relevant data and terminology standards. As the study is a randomized controlled trial, we will include the completed CONSORT flow diagram and checklist. For the cortisol biomarker datasets, we will follow guidelines for reporting of steroid hormonal assay measurements according to the guidelines adopted by the Endocrine Society. In addition, reporting of assay measurements for both the cortisol and inflammatory markers will include the name of the commercial kit used for assay, reports of the sensitivity of the technique and the within- and between-assay coefficients of variation for the assays as conducted in the laboratory. Information on the number of specimens lower than the minimum detection level of the assay or above the standard curve of the assay will be reported as will any procedure used to account for this data along with the rationale. For the urine drug toxicology screening (point of care screening with results immediately available), we will include the name, sensitivity and standards of the commercial kit used for screening.

**4. Data preservation, access, and timelines**

Scientific data will be shared publicly through the NAHDAP no later than the time of publications addressing the study aims or the end of the performance period, whichever comes first. Information on where to access data associated with a publication will be included on all publications. Electronic copies of publications will be deposited within four weeks of acceptance by a journal in PubMed Central with proper metadata to be made discoverable and accessible upon publication. The mission of the NAHDAP is to preserve deposited files in perpetuity.

**5. Access, distribution or reuse considerations**

*A. Factors affecting subsequent access, distribution, or reuse of scientific data:*

Study respondents will be asked to assent/consent to data collection and sharing deidentified data with the wider research community. If a minor participant turns 18 years old during the study, we will review the adult consent form with them, and if they consent to continued participation, we will have them sign the form. However, the informed consent/assent forms will be reviewed and approved by the OSU IRB to ensure data may be shared as proposed.

*B. Control of access to scientific data*:

As noted, deidentified data will be deposited and made publicly available through the NAHDAP, which is supported by the National Institute on Drug Abuse as a data repository. Users of the public use data must register with NAHDAP and agree to the Terms of Use. Deidentification will be completed by the end of data processing, prior to the finalization of the public use data files.

*C. Protections for privacy, rights, and confidentiality of human research participants*:

Identifying and personal health information will NOT be shared to protect participant confidentiality, including data related to birthdates, names and GPS coordinates of youth routine locations (only characteristics of these locations will be made available), and names and contact information of network members.

**6. Oversight**

The Pls of the proposal will discuss and make the DMS Plan available to all personnel involved in the project. The Pls will be responsible for ensuring faithful adherence to the DMS Plan, including review of the plan annually with the study team at a minimum with revision as needed as the research project evolves. Any potential revision will be discussed and approved beforehand with the NIH program official. Progress and compliance on this Plan will be communicated annually in the RPPR and in the final report.