

DATA MANAGEMENT AND SHARING PLAN

Element 1: Data Type

A. Types and amount of scientific data expected to be generated in the project:

The study will provide prospective, longitudinal data on the health outcomes during the post eviction moratorium (HOPE-M) period on our target sample of adults at risk for the threat of eviction. Quantitative and qualitative data will be collected. There will be 4 sources of quantitative data: court reported data on eviction filed status, census tract-level housing precarity risk estimates, COVID-19 county-level administrative data on supportive policies via the US COVID-19 County Policy data set, and self-reported survey data on economic hardship, psychosocial stressors, mental health, and socio-demographic characteristics.

Eviction status, census-tract aggregated housing precarity risk estimates, COVID-19 county-level administrative data, and self-reported data will be collected on 2000 participants. The greatest amount of data will be generated from the self-reported data. The self-reported data will be collected via surveys from 2000 participants at 3 time points (baseline, 6 months, 12 months) for a total of 6,000 surveys nested within 2000 participants. The survey data is being collected by M. Davis and Company, a survey evaluation firm. M. Davis will de-identify all data and merge the court reported data, the census tract aggregated housing precarity risk estimates, the county-level administrative data, and the self-reported survey data.

There will be 3 sources of qualitative data: focus group interviews, photographs and corresponding photo elicitation interviews. The research team will conduct 8 focus group interviews with a subset of 40 participants (5 participants per focus group group). The focus groups are estimated to last 90 minutes and will be audio recorded. Those same 40 participants will participate in one-on-one photo elicitation interviews. Each photo elicitation interview is estimated to last 60 minutes and will be audio recorded. Each participant will be asked to submit 6 photographs as a part of their photo elicitation interview for a total of 240 photographs nested within 40 participants. All focus group (n=8) and photo elicitation (n=40) audio recordings will be transcribed verbatim and de-identified into a word document. The interviews and photographs will have a linking code to the quantitative data through a participant identification number created by M.Davis and Company.

B. Scientific data that will be preserved and shared, and the rationale for doing so:

Quantitative and qualitative data will be stored in our secure computing environment at University of Texas Health Science Center at Houston (UTHealth). Any participant identifiers in the quantitative and qualitative data will be removed and maintained in a secure file that is only accessible by a few designative members of the research team. The rationale for preserving participant identification (name, email, phone numbers) is for the research team to be able to contact participants in the future. Future contact purposes include participation in supplemental projects or new projects that include the same population.

Data that is shared will be person-based and prepared for the broadest level of accessibility feasible through the research team's efforts of removing identifying information. The research team will consider secondary alternative variables and fields with less specificity (e.g., broader categories, age ranges rather than a specific age) to be sure that no person in the archived dataset can be identified.

C. Metadata, other relevant data, and associated documentation:

Documentation for the quantitative data will include a user's guide and a codebook. The user's guide will include a brief study overview, description of the eligibility criteria to be in the study, data collection procedures, and explanation of the variable structure. The user's guide will also list the topics in the data set. Underneath each topic the name of the survey instrument used to measure the topic will be included along with the corresponding citation(s). A list of variables in the dataset that correspond to each survey instrument will also be included. If composite variables need to be created, a short description will also be included. The user's guide will also include the funding source and the acknowledgement of NIH award support that needs to be included in each publication, press release, or other document,

The codebook will include detailed information about the variables in the data set. Each variable in the codebook will include the variable name, variable description (i.e. brief text from the survey question), and value labels (i.e. numeric codes and text description of the values). Documentation of the user's guide and codebook will be provided in PDF file.

Qualitative data (focus group interviews, photo elicitation one-on-one interviews and corresponding photos) will also be de-identified, with photo meta-data removed to protect participant privacy. De-identified raw transcripts of the focus group interviews, photographs, and photo elicitation interviews will be made available. The code needed to link the quantitative and qualitative data will also be provided.

Element 2: Related Tools, Software and/or Code:

Quantitative raw data will be released in excel, which is the most accessible format for various statistical software packages. Qualitative de-identified transcript (raw) data will be released as Word documents, and photos will be released after photo meta-data is removed.

Element 3: Standards:

Quantitative data will include variables that represent court-ordered eviction status, census tract aggregated housing precarity risk estimates, administrative data on COVID-19 related policies, and self-reported participant responses from standardized surveys focused on economic hardship (e.g., food insecurity), psychosocial stressors (e.g., perceived housing discrimination), mental health (e.g., anxiety, depression, suicide ideation), and socio-demographic characteristics (e.g., similar items to the Behavioral Risk Factor Surveillance System). It is intended that the user's guide and codebook be used jointly with the data.

Element 4: Data Preservation, Access, and Associated Timelines

A. Repository where scientific data and metadata will be archived:

The Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan has a self-publishing repository for social, behavioral, and health sciences research data called openICPSR. OpenICPSR will serve as the data repository for this study. openICPSR is particularly well-suited for the deposit of replication data sets for researchers who need to publish their raw data associated with a journal article so that other researchers can replicate their findings. ICPSR does not improve or alter datasets deposited in openICPSR in any way. Data are preserved as-is and distributed in the same condition and format submitted by the depositor.

B. How scientific data will be findable and identifiable:

The Data Management and Sharing Plan will be stored in UTHealth's Institutional Data Ecosystem Portal (DEPUT). The DEPUT will include direct links to where the data files and corresponding materials can be found.

C. When and how long the scientific data will be made available:

Research data will be available upon publication of related work or 6 months from the end of the project period, and will remain available as long as required by institutional policy or the sponsor.

Element 5: Access, Distribution, or Reuse Considerations

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

No additional limitations other than the controls and privacy protections described below.

B. Whether access to scientific data will be controlled:

Due to the identifying nature of the data collection process the following data will not be share as it could compromise privacy and confidentiality: court filing data, court captions, address information, including dates of residency, owner/renter status, and geospatial assessments. Photographs that directly identify individuals, identify where they may live, or could be used to identify participants will not be released.

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

After merging all four sources of quantitative data, all personal identifiers will be removed and maintained in a separate control file by M. Davis Company for future contact purposes. Future contact purposes include participation in qualitative interviews and/or any supplemental projects. Personal information in the qualitative data will be removed from all transcripts to maintain confidentiality.

Element 6: Oversight of Data Management and Sharing:

The Office of the Executive Vice President & Chief Academic Officer (EVP/CAO) and The Office of Data Science (ODS) at UTHealth Houston will provide joint institutional oversight for the Data Management and Sharing (DMS) plan. Datasets resulting from this research will be cataloged within the institutional DEPUT system. DEPUT is the institutional oversight management portal supported by UTHealth Houston for DMS validation and tracking. Project Contact PI (Dr. Daphne Hernandez) will update data status in DEPUT, and the institutional office of Sponsored Projects Administration will perform annual validation according to the DMS plan. Validation results will be reported to EVP/CAO and ODS for review. Gaps, if any, will be identified with appropriate correcting measures implemented. The PI will have overall responsibility for compliance with data collection, storage, and safety protocols.