**Study:** Perception of Water Insecurity and Epigenetic Signatures of Stress Among Urban Adult Women in Mexico City

**Study overview:** This research project investigates the relationship between perceived water insecurity, psychological stress, and biological markers of stress among adult women in Mexico City. The study focuses on understanding the impact of water access challenges, emotional responses to water insecurity, and potential biological stress indicators like DNA methylation.

**Location**: Iztapalapa, Mexico City, Mexico

**Target Population:** Adult women (self-identified), ages 18 to 45, living in neighborhoods with varying levels of water insecurity in Iztapalapa. Pregnant and lactating were excluded from the study.

**Sampling Method**: Opportunistic sampling, included a total of 397 individuals

**Data collection period**

* Season 1: Oct-Dec 2022
* Season 2: Apr-Jun 2023

**Data collection method**

* Surveys were paper-based and applied through in-person interviews, conducted by one of the PIs (Contreras) and three Mexican research assistants (two biologists and two anthropologists).
* Data collection took <50 minutes per person
* Participants were invited to participate in the study near neighborhood community centers where many women exercise, take classes, or bring their children for enrichment activities.
* All individuals provided IRB consent ((HUM00192307) and received economic compensation.
* Completed surveys were anonymized and coded into a secure database.
* Data were collected through:
  + Screening survey
  + NSE-AMAI
  + HWISE
  + PSS
  + QUAIL-Mex

**Data Preparation Notes**

* Across all surveys, a set of suffixes were used:
  + \_INFO: any additional notes written by the surveyors
  + \_CAT: information extracted from open-ended questions into categorical variables.
  + \_SC: score
  + \_TOTAL: value calculated from other variables included in the dataset
  + \_WI\_: stands for water insecurity
* Participant ID numbers go from 1 to 204 (season 1) and then from 301 to 497 (season 2). **I skipped numbers 205 to 300** to make clear that samples were collected at different times. ID number 164 was skipped.

**Missing information:**

* Empty cells represent missing information.
* Some questions have an important portion of the answers missing.
  + In those cases, the questions may have been added later to the survey, or we stopped asking them if they stopped making sense.
* Please code all missing info as N/A (no response).
* DK means “does not know”.
* There might be some answers with “pending” or the letter “P”, which stands for “Pending”. This info needs to be corroborated by looking at the original, hard-copy surveys.
* There are some participant IDs that do not exist, because they were removed or skipped. One of these IDs is ID-164.
* 493 should not be considered, as it does not meet inclusion criteria

**Instrument: SCREENING**

* Sociodemographic information (e.g., age, zip code, household size, health status)
* Some answers have been transformed into categorical data, but original data was kept in a separate column
* Variable names can have the following prefixes:
  + D (Demographics)
  + HLTH (Health)
  + W (Water-related)

**Instrument: NSE-AMAI**

* Stands for “Niveles Socioeconómicos-Asociación Mexicana de Agencias de Inteligencia de Mercado y Opinión”
* Variable names have the SES prefix
* Six-item tool developed specifically for Mexico

**Instrument: HWISE**

* Stands for Household Water Insecurity Scale ([Young et al. 2019](https://gh.bmj.com/content/4/5/e001750))
* 12-items
* Likert scales (4 options)

**Instrument: PSS**

* Stands for Perceived Stress Scores ([Cohen 1994](https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft02889-000))

**Instrument: QUAIL-Mex**

* Questionnaire focused on water access, storage, treatment, and emotional responses.
* Purpose:
* (A) identify if objective water insecurity is associated with perceived stress among women in Mexico City
* (B) identify the components of water insecurity that may be a source of distress for these women. This information will reveal if and how different aspects of water insecurity, such as unpredictable, unsafe, or limited water supply are consciously recognized as stressful for study participants, or alternatively, if they are accepted and normalized.
* It includes modified questions from the NESTSMX project ([Hubberts et al. 2023](https://journals.plos.org/water/article?id=10.1371/journal.pwat.0000056)) and other studies on water insecurity (Wutich, 2006).

In this study I applied 4 different surveys/questionnaires:

* **Screening**: age, neighborhood, health status, household characteristics. etc
* **SES**: instrument developed by a Mexican agency to measure socioeconomic status based on the current characteristics of the population. Each answer has a specific score and summing up all answers gives us a final score that is associated with a specific socioeconomic group (A/B, C+, C, C-, D+, D-, E)

|  |  |
| --- | --- |
| A/B | 205 or more |
| C+ | 166-204 |
| C | 135-165 |
| C- | 112 a 135 |
| D+ | **90-111** |
| D- | **48 a 89** |
| E | **less than 47** |

* **HWISE**: An internationally used 12-item questionnaire to measure water insecurity. Higher scores mean higher water insecurity. Some people calculate subscores from this survey, by selecting a few questions and adding up their scores.
  + Water worry subscores
* Questions used: worry, anger, shame
* Hygiene subsores
  + Questions used: clothes, hands, body
* **PSS:** Perceived stress scores, an internationally used scale to measure perceived stress. Scores are calculated by adding up questions 1,2,3, 8, 11, 12, 14, and subtracting 4,5,6,7,9,10, and 13
* **MEX questionnaire:** a set of questions selected and modified by me to better capture the local reality in Mexico City. All files that begin with “Q” contain answers for questions included in this questionnaire. Most questions were open ended, and I am in the process of transforming some answers into categorical variables.

Merging:

* Some observations were entered twice to test inter-coding error
  + Please flag IDs that present inconsistencies
* Some files have information for only one season: this is why it seems that some questions are in more than one file.

Variable names:

* They have a prefix, root and a suffix.
  + Prefix represent the group or type of question
    - e.g. all Health related questions start with “HLTH\_”
  + Root represents what the question is about
    - E.g. all information related to location or neighborhood have the root “LOC”
  + Suffixes indicate if a variable is categorical (CAT), a score (SC), or if it contains comments or other types of info (INFO). Suffixes are not present in all variables.