

# SSP Point in Time Survey Mock

## Data Analysis Plan

### Data collection:

Data collection will take place from XX/XX/202X - XX/XX/202X. Before administering the survey, verbal affirmative informed consent will be required for participation. More detailed information is stored in the Point in Time Survey Protocol.

The survey instrument collects information on demographics, as well as drug use, and harm reduction practices. A full version of the survey can be found at the end of the analysis plan and a brief version can be found in the table below.

Table 1. Variable names, questions as well as collection levels for data collected in the survey.

Variable	Question	Levels
Age	How old are you?	Age in years
Sex assigned at birth	What was your sex assigned at birth?	Radio required 1 Male 2 Female 3 Intersex 4 Declined to Answer
Gender	What best describes your gender identity?	checkbox, Required 1 org_pits_gender__1 Man 2 org_pits_gender__2 Woman 3 org_pits_gender__3 Trans Man 4 org_pits_gender__4 Trans Woman 5 org_pits_gender__5 Non-Binary 6 org_pits_gender__6 Two-Spirited 7 org_pits_gender__7 Other 8 org_pits_gender__8 Declined to Answer
Race/ethnicity	What race/ethnicity are you?	checkbox, Required 1 org_pits_ethnicity2__1 American Indian/Alaska Native 2 org_pits_ethnicity2__2 Asian/South Asian 3 org_pits_ethnicity2__3 Black/African American 4 org_pits_ethnicity2__4 Latinx/Hispanic 5 org_pits_ethnicity2__5 Native Hawaiian/Pacific Islander 6 org_pits_ethnicity2__6 White 7 org_pits_ethnicity2__7 Other 8 org_pits_ethnicity2__8 Declined to Answer
Ethnicity	What is your ethnicity?	Radio, Required 1 Hispanic/Latinx 2 Not Hispanic/Latinx 3 Don't know 4 Declined to Answer
Housing	How would you describe your housing?	Radio, Required 1 Temporary or unstable 2 Homeless 3 Permanent 4 Declined to Answer
Drug use	Which of these drugs have you used in the	checkbox, Required 1 org_pits_druguse6m__1 Speedball (heroin mixed w/ cocaine) 2 org_pits_druguse6m__2 Goofball (heroin mixed w/ meth)



	last six months? Check all that apply	3 org_pits_druguse6m__3 Heroin on its own 4 org_pits_druguse6m__4 Meth on its own 5 org_pits_druguse6m__5 Powder Cocaine on its own 6 org_pits_druguse6m__6 Crack Cocaine 7 org_pits_druguse6m__7 Fentanyl 8 org_pits_druguse6m__8 Prescription Opiates 9 org_pits_druguse6m__9 Other 10 org_pits_druguse6m__10 Unsure 11 org_pits_druguse6m__11 Declined to Answer
Injection drug use	Have you injected drugs in the last six months?	radio, Required 1 Yes 2 No 3 Unsure 4 Decline to Answer
Incarceration	Were you in a jail, prison, or correctional facility in the last 12 months?	radio, Required 1 Yes 2 No 3 Declined to Answer
Transactional sex	In the last six months, have you traded sex for money, drugs, housing, or another need?	radio, Required 1 Yes 2 No 3 Unsure 4 Decline to Answer

We have a data use agreement that provides our harm reduction organization with ultimate ownership of the data. This agreement ensures that if anyone would like to use the collected data it must be approved by us, and have us included in the work in the capacity of our choosing.

**Data cleaning:**

The data cleaning process will be undertaken in Microsoft Excel. Any cleaning or editing of the data will be done in a copy of the file with the raw data file always remaining intact.

- The copy version of the file will be denoted with the file name “data-pit-cleaning-copy.csv” and will be distinguished from the original raw data file, which will be named “data-pit-raw.csv”.
- We will look at collected data for spelling errors and correcting those so that there is uniformity in all the answers in accordance with how they were set out in the table above and the survey codebook.
- We will also look at if all the values fit within plausible ranges including: age, number of overdoses, number of witnessed overdoses, and number of ER visits. If values are not plausible we will first follow up with the person who administered the survey to ask if they remember and can identify if it was a data entry error. If the data cannot be resolved in this manner than the value in question will be set to NA.

*Checking unique IDs:*

Unique IDs will be verified for any duplicates using conditional formatting in Excel using the duplicate condition. We will then sort the data based on the unique ID column which will allow for the easy identification of duplicate unique IDs. Duplicate unique IDs will be checked to see if they are exact duplicates, which would result in further investigation to make sure the same survey was not entered twice.

### *Dealing with matching unique IDs:*

Ideally everyone in the dataset would be unique and only participated in the PIT survey once. Therefore, everyone should have a unique ID; however, there is possibility for duplication, where either one person has taken the survey twice or two (or more) people have the same unique ID.

In the event that there are multiple individuals with the same unique ID we will first make sure that the records are indeed distinct from one another by looking at the responses to questions of race, gender, and drug use to see if they are the same. If the responses are the same we will cross check other question responses to compare for similarities. We will also talk to the interviewer(s) who administered the survey for more details. If the answers to the questions are exactly identical and the interviewer(s) thinks they are the same person one of the observations will be removed from the analytical dataset, but kept in the original dataset from collection. A note will be made by the researchers that a record was removed in the report methods.

In the event that these constitute different individuals a two-step process will be done to create new unique IDs. First, a new column called "unique-id-2" will be made with the exact same values as the unique ID column. Second, we will sort the spreadsheet by unique ID and then add a number to the end of the non-unique ID starting at 1 and moving up by 1 integer per non-unique ID until all IDs are unique. For example, three individuals who have the unique ID AP94 will have the values "AP941, AP942, and AP943" for the variable "unique-id-2".

### **Data analysis:**

After the survey concludes, the data manager and lead analyst will collaborate on the data analysis. Numerical variables will be analyzed using Excel by looking at the range, a count of how many instances of each number occurs, as well as the mean and median and corresponding measures of spread standard error, and interquartile range. We will also be calculating descriptive statistics by looking at the distribution of categorical variables as counts. Exceptions for how categorical variables will be handled are for the race and gender variables. How we plan to analyse these two important variables are described in more detail below.

### *Calculating race:*

Race will be calculated through simple sums, but will also account for individuals with multiple racial identities in a new option called multiracial. This will be done through creating a new variable called, "race-analysis" to be used for the analysis portion and will consider the select all that apply answers over the X number of race variables that are part of the select all that apply for the race question. The new variable "race-analysis" will be created by looking at the responses to all the select all that apply race questions and then entering into the new variable column the corresponding race if there is one response and will enter "multi-racial" if more than one race is checked off in the original race variables.

### *Calculating gender:*

Gender will be calculated through a simple sum, but will be done on a new variable called "gender-analysis". The "gender-analysis" variable will be based off a combination of the sex assigned at birth variable as well as the gender variable. Individuals who have the same sex assigned at birth and gender will be given the value that they answer for the gender variable for "gender-analysis" with the prefix "cis-" added to the variable name. Those who identify within the gender binary, but have a different value for sex assigned at birth and the gender variable will be given the value they answer for the gender variable for "gender-analysis" with the prefix "trans-" added to the variable. Those who select a

gender that is not in the gender binary will be given the gender for “gender-analysis” as indicated in the gender question.

*Stratified analysis:*

We will also look at the demographic and other variables stratified by key variables of interest like gender, housing, and injection drug use in the last 6 months. This will allow us to compare the differences in demographics and risk behaviours among different segments of our clientele allowing for us to better adapt and tailor our services.

