

# Team 04: Estimating the Number of Intravenous Drug Users in Philadelphia

## Executive Summary

- Indirect estimation methods are essential tools for estimating the number of intravenous drug users (IDUs) and opioid users in a population.
- Addition estimates enable the use of multiple, publicly available data sources, but there is a risk of double-counting a population member.
- Multiplication method estimates enable the use of a single dataset, but these estimates depend on the accuracy of the multiplier used.
- Capture-Recapture estimates can be applied to Prevention Point's data recording needle exchange visits by using a Truncated Poisson Estimator, but the estimated population size varies significantly based on the interval of time considered for the data.

## Contributors

- Cara Cuiule
- Kelsey Keith
- Spandana Makeneni
- Sam May

## Problem definition and dataset

Our overall question was, "How many intravenous drug users are in the City of Philadelphia?" Data pulled from the City of Philadelphia precluded any methods that relied on granular data detailing a single individual. Data pulled from the City of Philadelphia included on fatal and nonfatal overdoses, medically assisted treatment, and drug arrests. Using these data required indirect estimate methods that relied on figures pulled from the literature to complete the indirect estimates.

Data recording needle exchange visits were provided by Prevention Point. These data enabled us to count distinct individuals who visited a needle exchange site during 2019. Using the needle exchange data, we attempted to estimate the number of intravenous drug users (IDUs) who possibly would have visited a needle exchange site but did not in 2019. By estimating the number of IDUs who attended the needle exchange 0 times, we are able to estimate the total number of IDUs in Philadelphia.

## Data

### Data for Addition, Multiplier Estimates

### Arrests

OpenDataPhilly is a collaborative, open-source project that provides access to datasets about the Philadelphia region, including data on arrests. Arrest data comes from the Philadelphia District Attorney's office and OpenDataPhilly provides a table of daily arrest counts that was summarized on a yearly level for this analysis.

## Overdoses

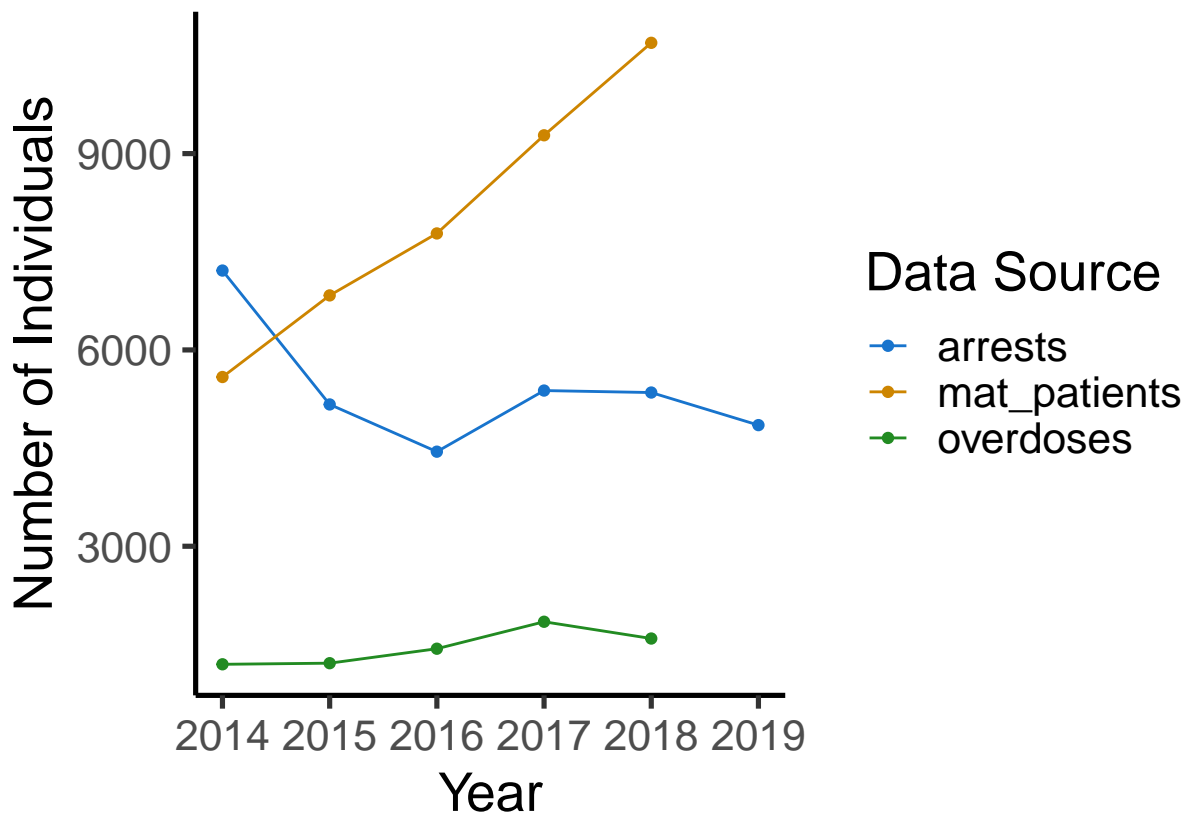
From Philadelphia Department of Health's public opioid tableau page <https://public.tableau.com/profile/pdph#!/>. There are two sources: fatal and nonfatal overdose deaths.

## Individuals in Medication Assisted Treatment (MAT)

From Philadelphia Department of Health's public opioid tableau page <https://public.tableau.com/profile/pdph#!/>. The Department of Health has data for individuals receiving Medicaid who are also receiving medication assisted treatment for opioid use disorder and for individuals in Philadelphia jails receiving MAT.

## Visualize the Data

Visualizing the data used to construct the estimates across the years we were able to find data for. Individuals experiencing an overdose and individuals receiving medication assisted treatment continue to increase, while arrests have generally decreased since 2014 and are now flat or slightly decreasing.



## Data for Poisson Estimate

Prevention Point provided a table of anonymized individuals who exchanged needles during 2019 that will be used for the poisson estimate. NOTE: For anyone who wishes to run the code underlying this report, the file with this data is password-protected and you will need to contact Code for Philly in order to access it.

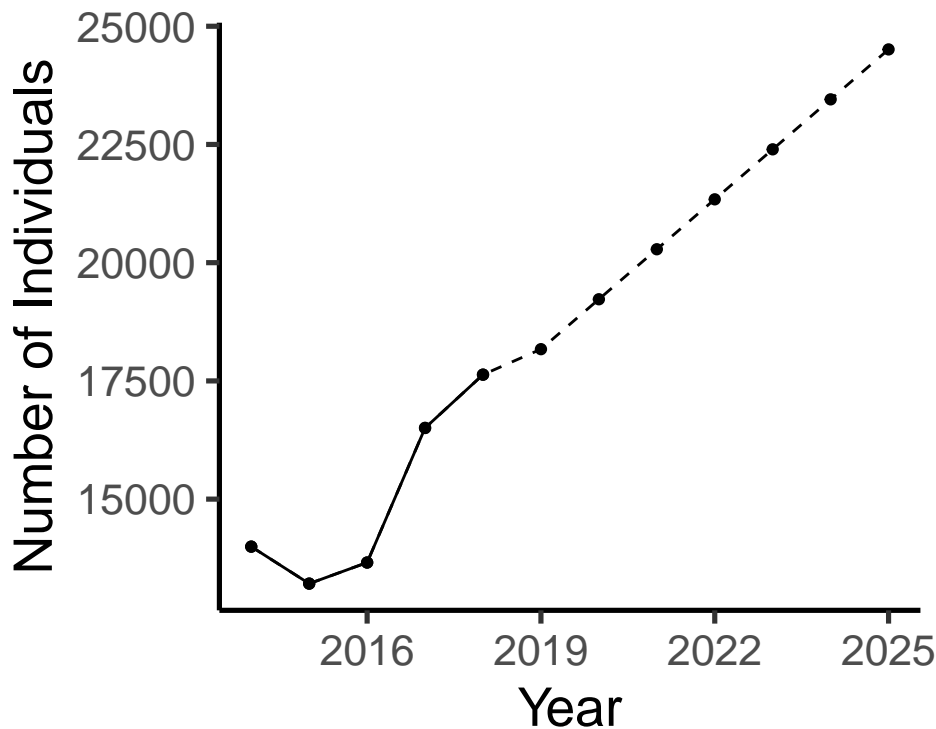
## Results

### Addition Estimate

For a lower-bound estimate of the number of opioid drug users (IDUs) in the city, we simply added the number of individuals arrested for drug possession + the number of overdoses presumably due to opioids + the number of people receiving medication assisted treatment for opioid use disorder. “Opioid drug user” is defined as an individual using heroin, fentanyl or fentanyl analogues. We didn’t find any reliable data for prescription opioid users, so they are not represented here. It was easier to first estimate the number of opioid drug users, then adjust the number down to an estimate of an estimate for intravenous drug users.

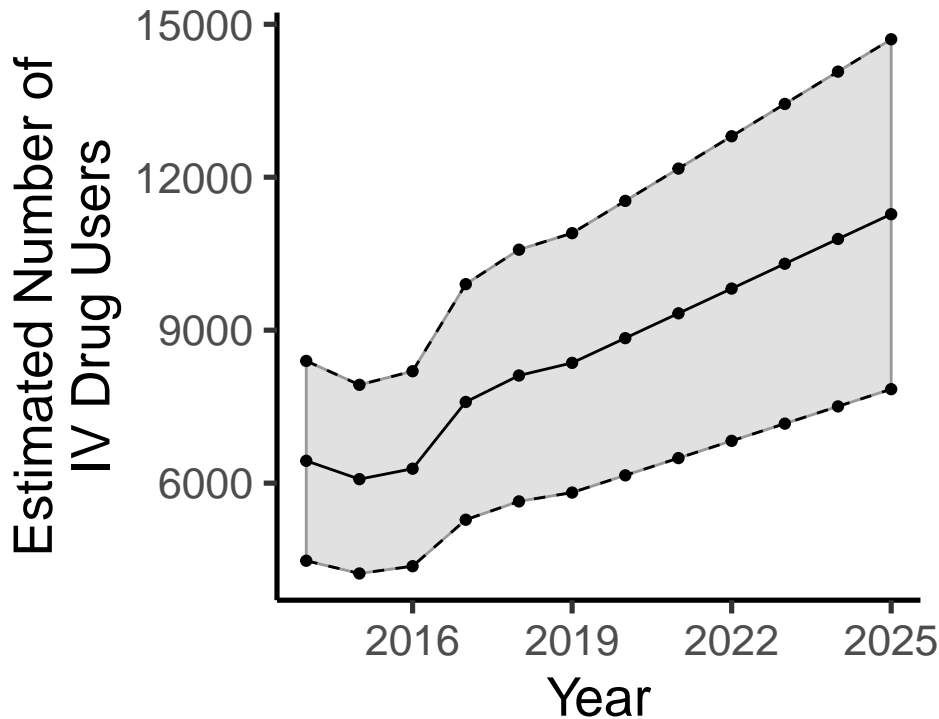
### Estimated Number of Opioid Users

Plot estimated number of opioid users. Note, we mostly were unable to find good data for 2019, so 2019 is estimated as well. The future project is linear, meaning we expect the number of opioid users to increase consistently year over year. However, there wasn’t sufficient data to fit the shape of the distribution, so a linear projection was picked as the simplest case.



### Adjust Addition Estimate for Intravenous Drug Users

Now, we take the estimate for the number of opioid users and adjust it to 30-60% of the total. Based on our research (Data: 1,5), this is the estimated proportion of opioid users who report that they have injected drugs. As you can see in the plot below, we are NOT confident of this estimate, with our confidence interval ranging from 3,000 to 6,000 people over the course of the plot.



### Multiplier Method using Cincinnati Paper

A multiplier estimate takes a known percentage of individuals out of a larger who do something attempts to estimate the size of the larger group based on the subgroup. Because it is extremely difficult to know what percentage of people are engaging in a illegal activity, multipliers are usually estimated based on a survey of the scientific literature. Since we did not have the time or ability to conduct such a survey, we took multipliers from the paper at this link <https://jheor.org/article/9729-estimating-the-prevalence-of-opioid-use-disorder-in-the-cincinnati-region-using-probabilistic-multiplier-methods-and-model-averaging> looking at opioid users in Cincinnati in 2019.

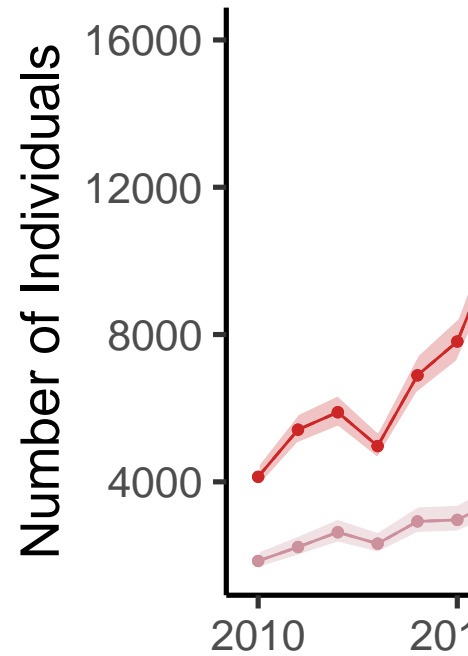
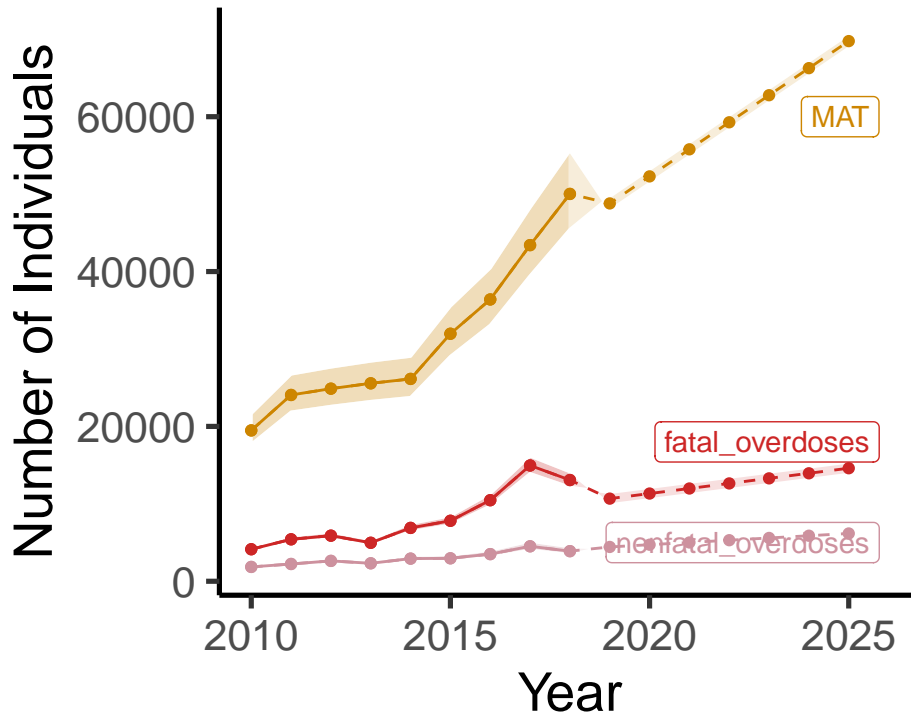
### Multipliers

- Fatal Overdose
- Multiplier: 7.182 %
- Std. Error: 0.533 %
- Treatment Admission
- Multiplier: 21.376 %
- Std. Error: 2.142 %
- Non-fatal Emergency Department (ED) Visit Overdose
- Multiplier: 40.890 %
- Std. Error: 5.111 %

### Multiplier Estimations of the Number of Opioid Users in Philadelphia

The lines on the plot below are colored and labelled by the data used to CALCULATE the estimate; all lines are attempting to estimate the number of opioid users in Philadelphia. The estimate based on MAT patients is wildly divergent from the rest of our estimates and is most likely wildly inaccurate. This is probably

because we and the Cinninati paper have different definitions of an MAT patient and/or access to MAT is different in Cinncanti. Don't use the MAT estimate. The overdose estimates are better.

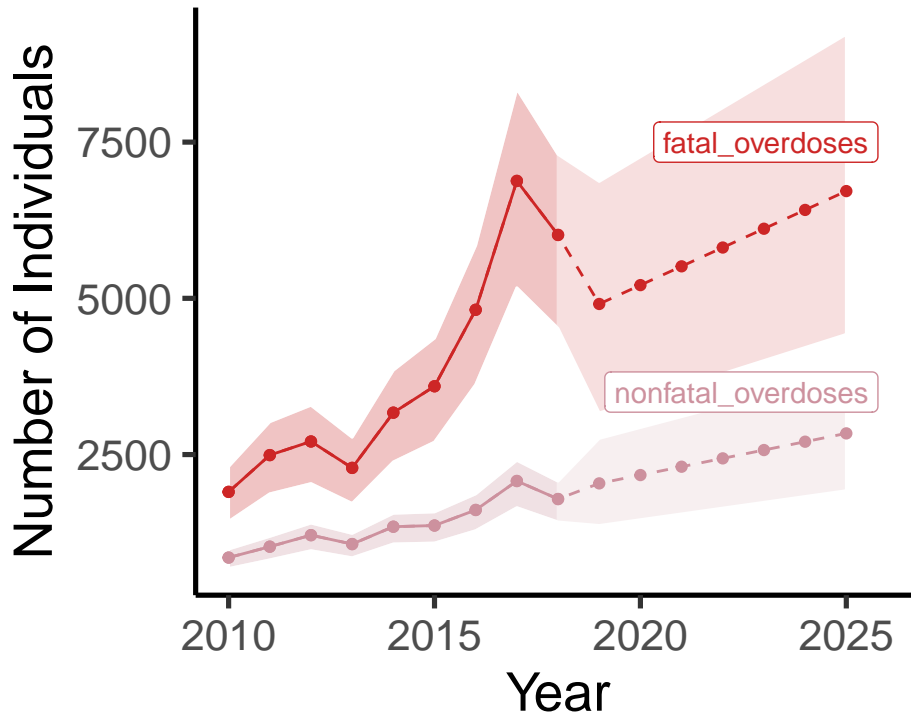


#### Adjusted Multiplier Estimates for IDUs

##### Adjusted Estimate

Again, we'll adjust the estimated number of opioid users to 30-60% of the total to estimate the number of intravenous drug users.

Because the estimate based on MAT patients was so divergent from all the other estimations in this report, we didn't adjust it down for the attempt to estimate the number of intravenous drug users. Again, we have wide confidence intervals on these estimates.



### Poisson Estimation Method

Prevention Point (PP) records each instance an individual visits the needle exchange. The data provided indicate a distinct individual with an anonymized unique identifier. The unique identifiers make it possible to count the frequency of an individual’s visits to the needle exchange. Using the observed frequency of visits by each needle exchange participant, we can apply the Poisson distribution to estimate the number of intravenous drug users (Hay & Smit 2003).

The Truncated Poisson Estimator is motivated by the need to estimate the number of individuals who have visited the needle exchange 0 times but could possibly have visited the needle exchange in the observed time period. This method is particularly applied to a set of indirect estimation methods, referred to as “Capture-Recapture experiments,” in which researchers track the number of times they encounter distinct members of the population being studied. The relative frequencies of individuals who visit within a certain time period are assumed to have a Poisson distribution. That distribution is estimated in order to estimate the number of individuals never encountered in the experiment (Zelterman 1988).

In the case of intravenous drug users (IDUs) in the Philadelphia area, we count the number of times a distinct participant visited a site in PP’s needle exchange program. We intuit that needle exchange participants who have only visited once or twice within the observed time period have many characteristics in common with intravenous drug users who have visited 0 times within the observed time period. The observed counts of distinct individuals who visited the needle exchange once or twice within a given time period enable us to estimate those who visited 0 times, giving us an estimate for the total number of IDUs in the Philadelphia area.

A key assumption of the Truncated Poisson Estimator is that the population size fluctuates only in a negligible manner, meaning that the population size should remain relatively constant throughout the entire observed time period (Hickman & Taylor 2005; Hay & Smit 2003). As can be seen below, the estimated population size varies significantly based on the set of observations used to make such estimate; this variance based on the interval of time observed has occurred in other studies of this sort (Hay & Smit 2003). The below table indicates the estimate results based on the observed time periods used to create such estimate. The

## Relative Frequency of Visits

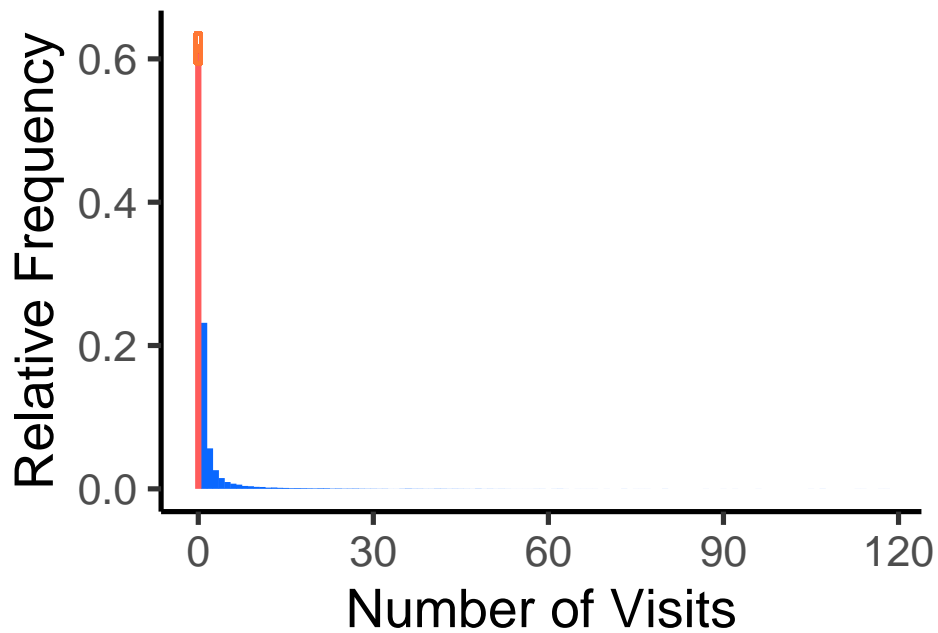


Figure 1: The relative frequency is calculated as the number of individuals who attended a certain number of times divided by the estimated size of the population. The blue frequencies indicate the individuals who were observed by Prevention Point within the data provided. The red bar indicates the estimated number of IDUs in Philadelphia who did not visit the needle exchange. The orange box at the top of the estimate reflects the 95% confidence interval of the population estimate based on all visits in 2019.

variation in the estimates suggests that it may be necessary to use shorter time periods, such as monthly data, instead of longer time periods, such as annual data in order to account for fluctuations in the size of the number of IDUs over time.

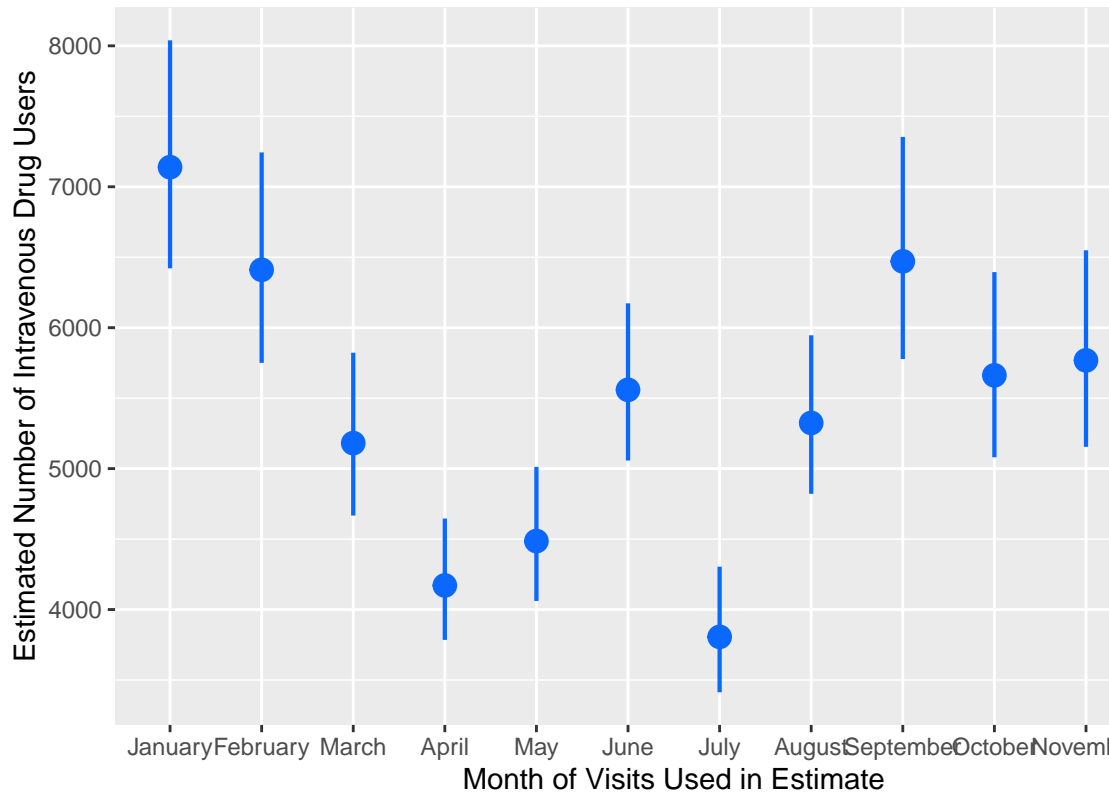
Table 1: Estimate of IDUs Based on Varying Observed 2019

	Estimated Population	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Total
All of 2019	28436	27852.397	29046.327	
Quarter 1	13843	13035.125	14757.889	
Quarter 2	9774	9285.698	10317.677	
Quarter 3	11529	10847.923	12303.328	
Quarter 4	12727	12021.187	13522.689	
January	7139	6421.084	8039.122	
February	6411	5750.399	7243.702	
March	5181	4666.907	5822.531	
April	4171	3784.700	4645.806	
May	4486	4060.832	5012.142	
June	5559	5057.399	6172.639	
July	3806	3412.725	4304.132	
August	5324	4821.156	5946.083	
September	6471	5777.791	7354.063	
October	5662	5080.345	6394.418	
November	5768	5154.564	6549.318	
December	7094	6384.624	7981.630	

The estimates based on monthly data can be displayed in a manner that helps indicate the uncertainty of the estimated population size. These estimates tend to overlap, indicating that the population size may be best



Estimate of Intravenous Drug Users Based on Monthly Visits



measure on a monthly basis.

We additionally investigated how the estimate is adjusted based on very short time intervals. All possible, consecutive 5-day intervals were investigated. A set of possible dates was determined by finding all of the weekdays during which the needle exchange had at least one visit to a site. Using the possible list of dates, an estimate was created for each possible consecutive 5-day period for the exchange’s operation.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1141	2195	2512	2513	2844	3644

Similarly, a random sample of 5 days was used to generate random samples of any 5 days during which the exchange operated. The resulting trials indicate the possible variation of an estimate created based on one week of operation by the PP needle exchange sites.

There is a significantly level of variation in the estimates generated based on a 5 days of operation by the needle exchange. We propose using monthly observations to continue making estimates in the future when applying the Truncated Poisson Estimator.

## Conclusions and Next Steps

These estimates are all bad, but the relative magnitude of the estimates in the low thousands are probably reasonable. The methods used here can be applied to better data either by us or by organizations within the city. Better, more granular data will allow us to make better estimates that avoid duplicatively counting the same person. More granular data would enable a more robust implementation of a Capture-Recapture experiment that could combine data collected at Prevention Point’s needle exchange sites with other data sources managed by the City of Philadelphia.

# Distribution of Population E

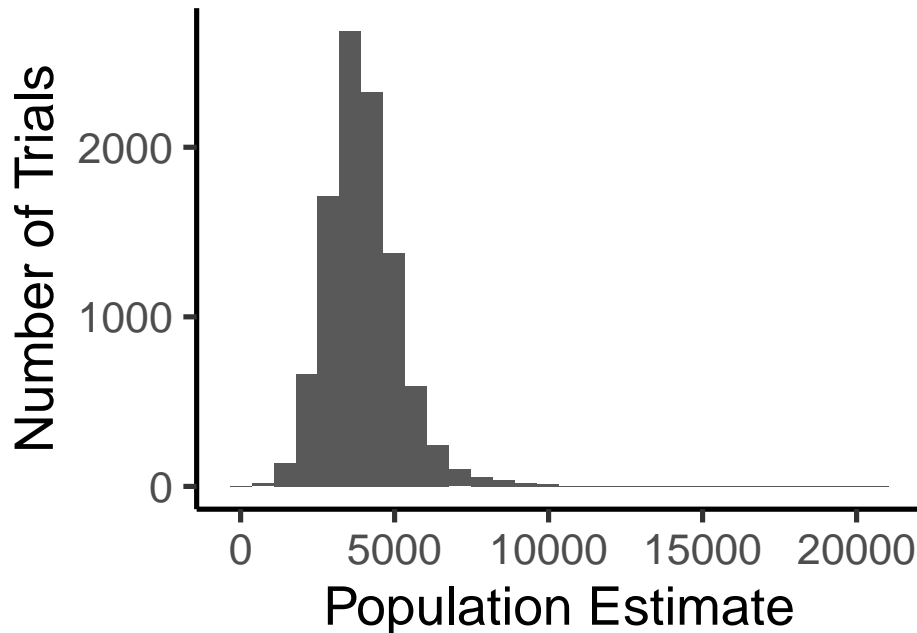


Figure 2: The figure indicates the distribution of the population estimate based on simulated 5-day intervals from 2019. 10,000 simulations are represented.

## References

### Literature:

1. Arnaud, Jeannin, and Dubois-Arber (2011) "Estimating national-level syringe availability to injecting drug users and injection coverage: Switzerland, 1996-2006," *International Journal of Drug Policy*, 22(3):226-232, doi: 10.1016/j.drugpo.2011.03.008
2. Gordon Hay & Filip Smit (2003) "Estimating the Number of Drug Injectors from Needle Exchange Data," *Addiction Research & Theory* 11:4, 235-243, DOI: 10.1080/1606635031000135622
3. Hickman, Matthew and Colin Taylor (2005) "Indirect Methods to Estimate Prevalence," *Epidemiology of Drug Abuse*, 113-132.
4. Mallow, et al (2019) "Estimating the Prevalence of Opioid Use Disorder in the Cincinnati Region Using Probabilistic Multiplier Methods and Model Averaging," *Methodology and Health Care Policy* Vol 6, Issue 2, <https://doi.org/10.36469/9729>
5. Zelterman, Daniel (1988) "Robust Estimation in Truncated Discrete Distributions with Application to Capture-Recapture Experiments," *Journal of Statistical Planning and Inference* 18, 225-237.

\*\*Data:\*\*

1. Levin, Jules (2018), "The Hepatitis C Continuum of Care for People Who Inject Drugs; Philadelphia, PA - young PWID 8% treated for HCV, older still only 25%" National Aids Treatment Project Conference Report [http://www.natap.org/2018/AASLD/AASLD\\_158.htm](http://www.natap.org/2018/AASLD/AASLD_158.htm)
2. Philadelphia Department of Public Health: <https://public.tableau.com/profile/pdph#!/>
3. Philadelphia Arrest Data from the District Attorney's Office via Philly Open Data: <https://www.opendataphilly.org/showcase/da-dashboard-arrests>

4. Prevention Point Needle Exchange Data: [https://github.com/CodeForPhilly/datahack2020/blob/master/data/pp\\_sep\\_site\\_event.zip](https://github.com/CodeForPhilly/datahack2020/blob/master/data/pp_sep_site_event.zip)
5. Treatment Episode Data Set(2010-2017): <https://datafiles.samhsa.gov/study-series/treatment-episode-data-set-admissions-teds-nid13518>