Technical and statistical issues in wastewater-based drug epidemiology

Jason R Williams

Alcohol & Drug Abuse Institute
University of Washington

11 March 2020

In which a social scientist walks into a wastewater treatment plant.

Outline

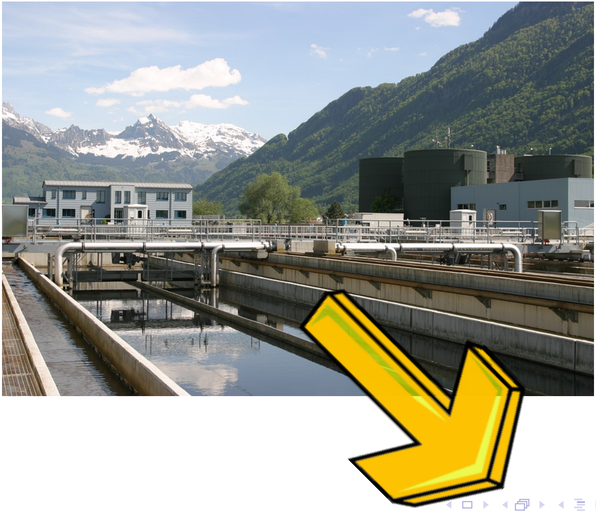
1. Wastewater-based drug epidemiology
   - WBE: Looking upstream to find hidden behaviors
2. From testing results to estimates
   - Assembling the pieces
   - Multiple pieces, multiple uncertainties
   - Estimation in the presence of censored data
   - Apply tools from other fields to generate a sample estimate
3. Visualization and results
   - Examples
4. Summary
   - Your take-aways

Outline

1. Wastewater-based drug epidemiology
   - WBE: Looking upstream to find hidden behaviors
2. From testing results to estimates
   - Assembling the pieces
   - Multiple pieces, multiple uncertainties
   - Estimation in the presence of censored data
   - Apply tools from other fields to generate a sample estimate
3. Visualization and results
   - Examples
4. Summary
   - Your take-aways

Wastewater-based Drug Epidemiology

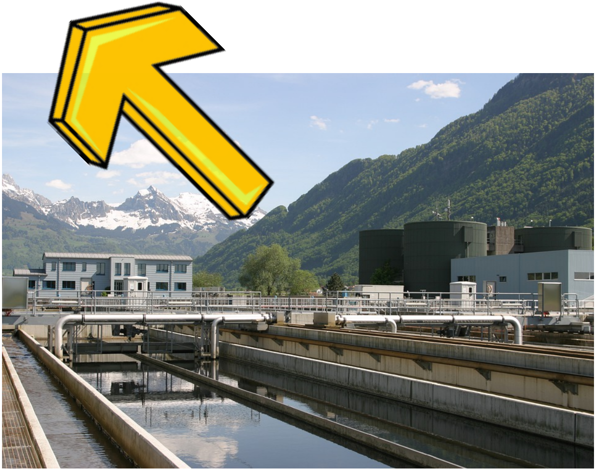
Looking upstream



- Wastewater treatment plants typically sample
    - Outflows
    - Inflows—to see how much of a pollutant is being removed

Wastewater-based Drug Epidemiology

Looking upstream



- Testing waste as it comes *in* to wastewater treatment plants (WWTPs) allows for quantification of:
  - Exposure to specific chemicals and foods
  - Markers of oxidative stress and allergic reactions
  - Levels of drug use

Wastewater-based Drug Epidemiology

Looking upstream

- Using wastewater to identify levels of drug use
  - Testing waste as it comes *in* to wastewater treatment plants
  - Requires assembling a number of pieces, drawing on analytic chemistry, environmental statistics, population estimation
  - Project involved 6 drugs, 19 WWTPs in Oregon and Washington
    - Banta-Green, C.J., Brewer, A.J., Ort, C., Helsel, D.R., Williams, J.R., & Field, J.A. (2016). Using wastewater-based epidemiology to estimate drug consumption—Statistical analyses and data presentation. *Science of the Total Environment, 568*, 856–863. `http://dx.doi.org/10.1016/j.scitotenv.2016.06.052`
  - Cannabis in wastewater project: 2 WWTPs serving Tacoma
    - Burgard, D. A., Williams, J., Westerman, D., Rushing, R., Carpenter, R., LaRock, A., Sadetsky, J., Clarke, J., Fryhle, H., Pellman, M., & Banta-Green, C. J. (2019). Using wastewater-based analysis to monitor the effects of legalized retail sales on cannabis consumption in Washington State, USA. *Addiction, 114*, 1582–1590. `https://doi.org/10.1111/add.14641`.

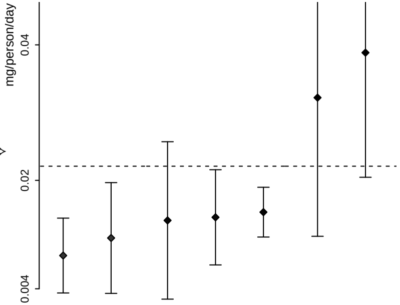
Outline

1. Wastewater-based drug epidemiology
   - WBE: Looking upstream to find hidden behaviors
2. From testing results to estimates
   - Assembling the pieces
   - Multiple pieces, multiple uncertainties
   - Estimation in the presence of censored data
   - Apply tools from other fields to generate a sample estimate
3. Visualization and results
   - Examples
4. Summary
   - Your take-aways

Wastewater testing results, substance X

City	Measured Conc. (ng/L)	Titration Correction Factor	Flow (L/day)
A	1005.2	1.077	17261424
A	1782.0	1.053	11318346
A	<LOD	1.056	15375268
A	<LOQ	1.049	15568495
A	1012.5	1.038	15456856
⋮	⋮	⋮	⋮

⟹



- Estimating yearly average per capita (or index) load

$$index\ load = \frac{\sum_{i=1}^{N} index\ load_i}{N} \tag{1}$$

$$index\ load_i = C_i \times F_i \div P_i \tag{2}$$

- *May* be able to multiply by a metabolization factor to estimate amount ingested

Wastewater-based Drug Epidemiology

Assembling the pieces

Wastewater-based Drug Epidemiology

Zuccato et al., 2005

"Our data suggest that actual cocaine consumption may be much greater than estimated by current methods....If we consider that in the River Po basin there are about 1.4 million young adults, the official [survey results] in this area would translate into at least 15,000 cocaine use events per month. We however found evidence of **about 40,000 doses per day**, a vastly larger estimate. The economic impact of trafficking such a large amount of cocaine would be staggering."

Wastewater-based Drug Epidemiology

Zuccato et al., 2005

"Our data suggest that actual cocaine consumption may be much greater than estimated by current methods....In agreement with these findings, cocaine loads determined at WWTPs gave drug consumption estimates of **about 2–7 doses per 1000 people, or 9–26 doses per day per 1000 young adults.**"

Wastewater-based Drug Epidemiology

Zuccato et al., 2005, Tables 1 & 2

	Levels[a]		Load	Use per 1000 people	
	Cocaine ng/liter	BE ng/liter	Cocaine[b] g/day	All g/day	Young adults[c] g/day
River Po	1.2 ± 0.2^e	25 ± 5^e	3800 ± 720^e	0.70 ± 0.13^e	2.7 ± 0.5^e
Cagliari	83	640	130	0.47	1.7
Cuneo	76	420	30	0.21	0.9
Latina	120	750	33	0.73	2.6
Varese	42	390	36	0.32	1.4
				0.44 ± 0.23^e	1.7 ± 0.7^e

[a]Cocaine and BE were analyzed by HPLC-MS/MS
[b]Cocaine-equivalent loads estimated from BE concentrations in the waters
[c]15–34 yr old
[e]Mean ± SD
River Po sampling: 4 days, 5 samples every 30 min each day
WWTP sampling: Single 24-hour composite, sampled every 20 min

Wastewater-based Drug Epidemiology

Banta-Green et al., 2009

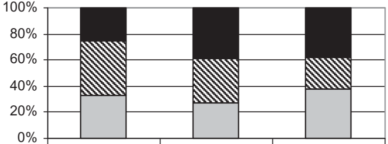
"The findings suggest a valid, rich data source that is complementary to other drug surveillance data sources.... Data were presented in terms of the relative distribution of index drug loads for each substance. This method of data presentation limits comparisons across drugs to whether substances were or were not detectable/quantifiable and **precludes direct comparisons of drug index loads.**"

Wastewater-based Drug Epidemiology

Banta-Green et al., 2009

"The findings suggest a valid, rich data source that is complementary to other drug surveillance data sources.... The **ongoing work of the study team is focused upon quantifying the uncertainty around computed index loads** and the source of index load variability to inform future sampling campaigns and analyses in order to make more refined comparisons between substances and locations."

Methamphetamine Level by Urbanicity

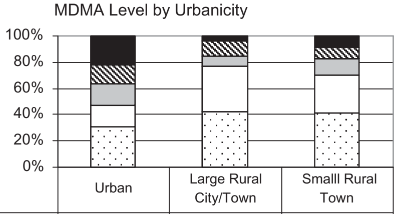


	Urban	Large Rural City/Town	Smalll Rural Town
■ Upper Tertile	9	10	13
▨ Middle Tertile	15	9	8
▦ Lowest Tertile	12	7	13
□ Below quantification	0	0	0
⊡ Not Detected	0	0	0

# of Municipalities

Wastewater-based Drug Epidemiology

Banta-Green et al., 2009



	Urban	Large Rural City/Town	Smalll Rural Town
Upper Tertile	8	1	3
Middle Tertile	5	3	3
Lowest Tertile	6	2	4
Below quantification	6	9	10
Not Detected	11	11	14

# of Municipalities

Figure 1 Number and proportion of single-day drug index loads by urbanicity in Oregon for benzoylecgonine (BZE) (cocaine metabolite), methamphetamine and 3,4-methylenedioxy-methamphetamine (MDMA)

Outline

- Using wastewater to identify levels of drug use
  1. Within-year sampling regime
  2. Within-day sampling at the WWTP
  3. Analytical chemistry (Large volume injection liquid chromatography/tandem mass spectrometry)
  4. Population estimate
  5. Excretion (metabolization) rates?
- Putting the pieces together: Complete estimates accounting for measurement error

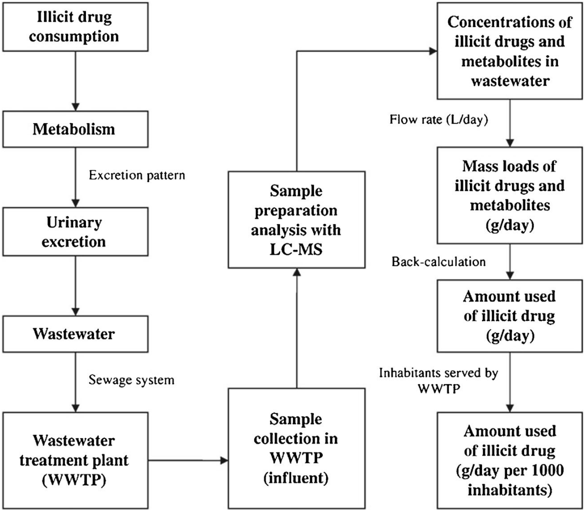
- Using wastewater to identify levels of drug use
  1. Within-year sampling regime
  2. Within-day sampling at the WWTP
  3. Analytical chemistry (Large volume injection liquid chromatography/tandem mass spectrometry)
  4. Population estimate
  5. Excretion (metabolization) rates?
- Putting the pieces together: Complete estimates accounting for measurement error

- Using wastewater to identify levels of drug use
  1. Within-year sampling regime
  2. Within-day sampling at the WWTP
  3. Analytical chemistry (Large volume injection liquid chromatography/tandem mass spectrometry)
  4. Population estimate
  5. Excretion (metabolization) rates?
- Putting the pieces together: Complete estimates accounting for measurement error

- Using wastewater to identify levels of drug use
  1. Within-year sampling regime
  2. Within-day sampling at the WWTP
  3. Analytical chemistry (Large volume injection liquid chromatography/tandem mass spectrometry)
  4. Population estimate
  5. Excretion (metabolization) rates?
- Putting the pieces together: Complete estimates accounting for measurement error

Wastewater-based Drug Epidemiology

Assembling the pieces



Meyer, 2014, *Anal Bioanal Chem*, Fig 3. Schematic overview of the sewage epidemiology approach

- Excretion rates are problematic.
  - In Burgard et al., 2019, we note that the two most prominent excretion estimates for THC to THC-COOH are 0.5% and 2.5%.
  - Like much cannabis research, these estimates are based on smoking.
  - Furthermore, the intermediate metabolite THC-OH may also be excreted, and itself metabolizes to THC-COOH in wastewater systems.

- Estimating yearly average per capita (or index) load

$$index\ load = \frac{\sum_{i=1}^{N} index\ load_i}{N} \tag{3}$$

$$index\ load_i = C_i \times T_i \times F_i \div P_i \tag{4}$$

$$index\ load_i = C_i \times T_i \times F_i \div P_i$$

- C_i is the estimated concentration in the sample
  - Possibly with censoring (indicated by unique codes to indicate $<$LOD or $<$LOQ)
  - Additional measurement error estimated by repeated testing 2 to 4 times
  - From LC/MS results

$$index\ load_i = C_i \times T_i \times F_i \div P_i$$

- T_i is a titration factor indicating how the sample was modified to facilitate chemical analysis
  - Assumed to be measured without error
  - Reflects storing and expanding the sample for testing
  - From LC/MS procedure

$$index\ load_i = C_i \times T_i \times F_i \div P_i$$

- F_i is the estimated flow of liquid into the WWTP for that day
  - Fairly established measurement
  - Use an estimated measurement error of 5% (RSD) from prior WWTP testing protocols (Brewer et al., 2012)

$$index\ load_i = C_i \times T_i \times F_i \div P_i$$

- P_i is the estimated population of users of the WWTP for that day
  - Ideally, want number of contributors to the flow
    - Toilet user-days?
  - In reality, have estimate of catchment area population from WWTP itself

$$\textit{index load}_i = C_i \times T_i \times F_i \div P_i$$

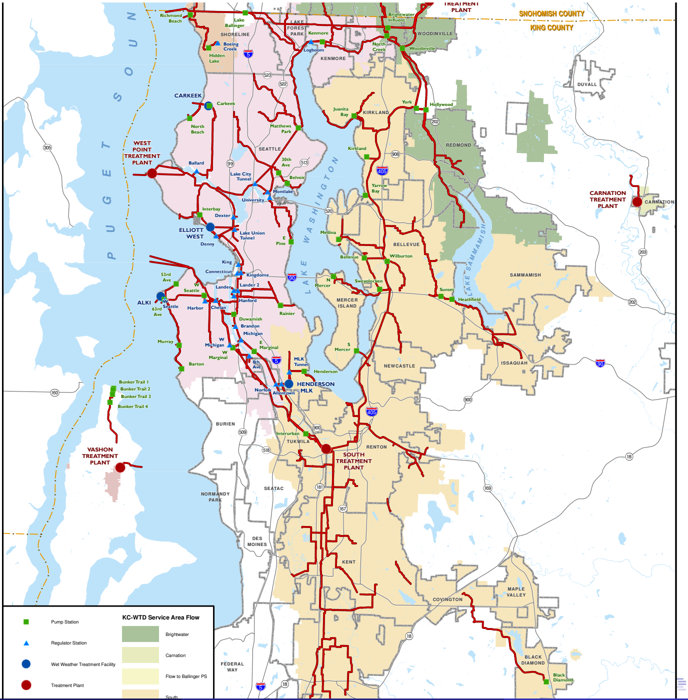
- P_i is the estimated population of users of the WWTP for that day
  - Ideally, want number of contributors to the flow
    - Toilet user-days?
  - In reality, have estimate of catchment area population from WWTP itself

$$\mathit{index\ load}_i = C_i \times T_i \times F_i \div P_i$$

- P_i is the estimated population of users of the WWTP for that day
  - We combine, where possible, Census Bureau-based daytime population correction and estimation error
  - Otherwise use prior error estimate of 20% (Ort et al., 2014)
  - Why "for that day"?
    - Daytime population correction is for weekdays.
    - Renton and Seattle WWTPs serve different catchments depending on rain and flow.

$$index\ load_i = C_i \times T_i \times F_i \div P_i$$

- P_i is the estimated population of users of the WWTP for that day
  - We combine, where possible, Census Bureau-based daytime population correction and estimation error
  - Otherwise use prior error estimate of 20% (Ort et al., 2014)
  - Why "for that day"?
    - Daytime population correction is for weekdays.
    - Renton and Seattle WWTPs serve different catchments depending on rain and flow.


SNOHOMISH COUNTY
KING COUNTY
TREATMENT PLANT
Brightwater Influent
Richmond Beach
Lake Ballinger
LAKE FOREST PARK
Kenmore
SHORELINE
Boeing Creek
Hidden Lake
Logboom
KENMORE
WOODINVILLE
North Creek
Woodinville
DUVALL
CARKEEK
Carkeek
North Beach
Matthews Park
Juanita Bay
KIRKLAND
York
Hollywood
REDMOND
WEST POINT TREATMENT PLANT
Ballard
SEATTLE
30th Ave
Kirkland
Lake City Tunnel
Belvoir
Montlake
University
Yarrow Bay
CARNATION TREATMENT PLANT
CARNATION
Interbay
Dexter
ELLIOTT WEST
Denny
Lake Union Tunnel
E Pine
Medina
BELLEVUE
Bellevue
Wilburton
King
Connecticut
Kingdome
53rd Ave
W Seattle
Lander
Lander 2
N Mercer
Sweyolocken
SAMMAMISH
ALKI
Seattle
63rd Ave
Harbor
Chelan
Hanford
Rainier
MERCER ISLAND
Sunset
Heathfield
Duwamish
Brandon
Michigan
Murray
W Michigan
E Marginal
W Marginal
S Mercer
ISSAQUAH
Barton
8th Ave
MLK Tunnel
Henderson
NEWCASTLE
Bunker Trail 1
Bunker Trail 2
Bunker Trail 3
Bunker Trail 4
Norfolk
Allentown
HENDERSON MLK
BURIEN
Interurban
TUKWILA
RENTON
VASHON TREATMENT PLANT
SOUTH TREATMENT PLANT
SEATAC
NORMANDY PARK
DES MOINES
KENT
MAPLE VALLEY
COVINGTON
BLACK DIAMOND
Black Diamond
FEDERAL WAY
PUGET SOUND
LAKE WASHINGTON
LAKE SAMMAMISH
Pump Station
Regulator Station
Wet Weather Treatment Facility
Treatment Plant
KC-WTD Service Area Flow
Brightwater
Carnation
Flow to Ballinger PS
South

Wastewater-based Drug Epidemiology

Assembling the pieces

- Combining uncertainties
  - Easier:

$$U_T = \sqrt{U_S^2 + U_C^2 + U_F^2 + U_P^2} \tag{5}$$

  - More advanced: Monte Carlo

Outline

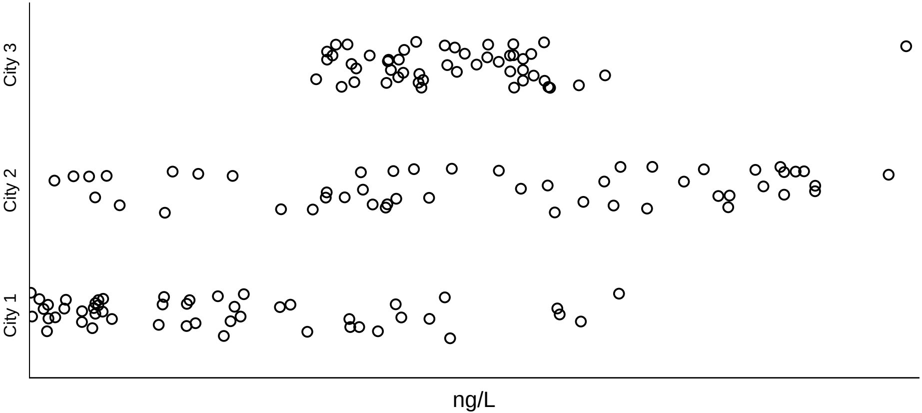
1. Wastewater-based drug epidemiology
   - WBE: Looking upstream to find hidden behaviors
2. From testing results to estimates
   - Assembling the pieces
   - Multiple pieces, multiple uncertainties
   - Estimation in the presence of censored data
   - Apply tools from other fields to generate a sample estimate
3. Visualization and results
   - Examples
4. Summary
   - Your take-aways

Wastewater testing results, substance X

City	Measured Conc. (ng/L)	Titration Correction Factor	Flow (L/day)
1	1005.2	1.077	17261424
1	1782.0	1.053	11318346
1	<LOD	1.056	15375268
1	<LOQ	1.049	15568495
1	1012.5	1.038	15456856
⋮	⋮	⋮	⋮

Censored measures

Multiple samples to generate more precise measure



Censored measures

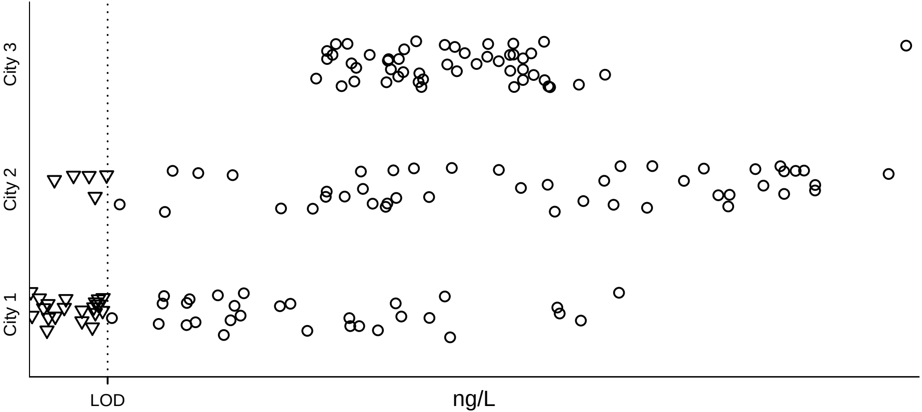
When measurement is less precise at lower levels

$<$LOD Below limit of detection ("non-detects")

- Left-censored: $x <$ LOD

Censored measures

Below limit of detection



Censored measures

When measurement is less precise at lower levels

$<$LOD Below limit of detection ("non-detects")

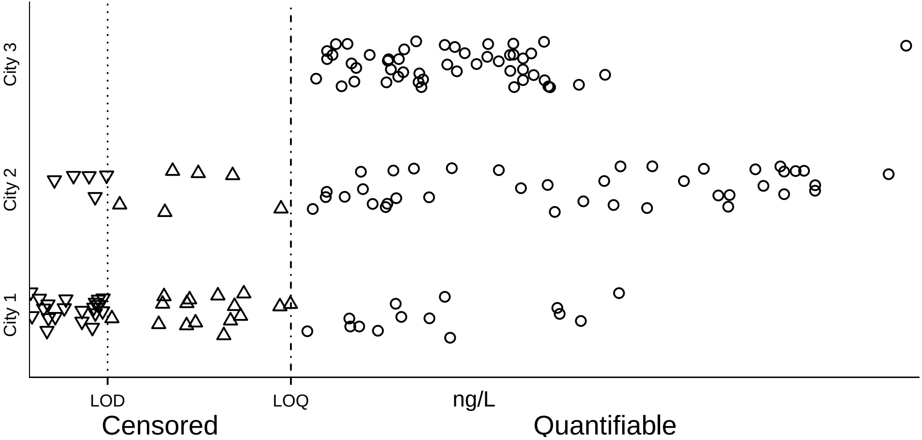
- Left-censored: $x <$ LOD

$<$LOQ Below limit of quantification

- Interval-censored: LOD $< x <$ LOQ

Censored measures

Below limit of quantification



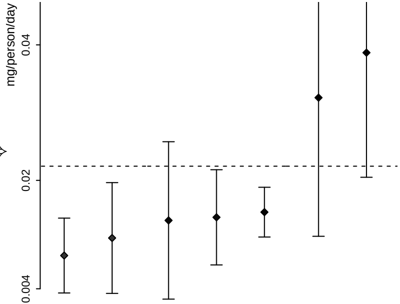
Estimation in the presence of censorship

How do we generate a mean?

Wastewater testing results, substance X

City	Measured Conc. (ng/L)	Titration Correction Factor	Flow (L/day)
A	1005.2	1.077	17261424
A	1782.0	1.053	11318346
A	<LOD	1.056	15375268
A	<LOQ	1.049	15568495
A	1012.5	1.038	15456856
⋮	⋮	⋮	⋮

⟹



Estimation in presence of censored data

Less robust solutions

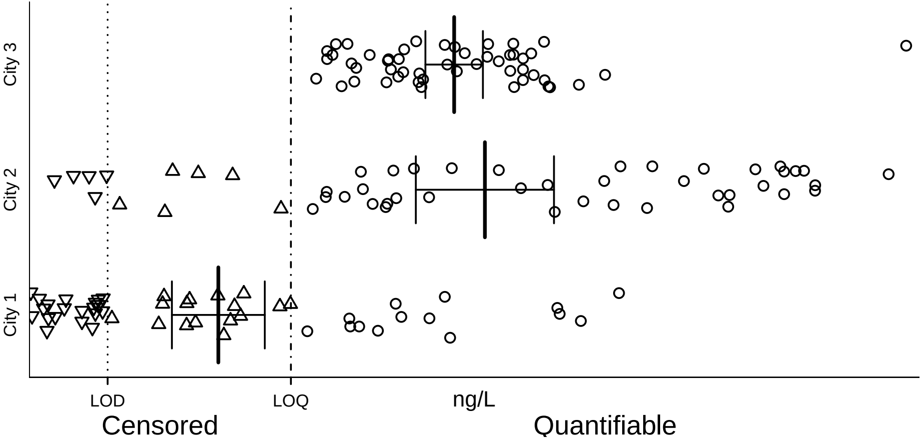
- Dealing with censored data: Ignore censored observations

  <LOD Below limit of detection → NA

  <LOQ Below limit of quantification → NA

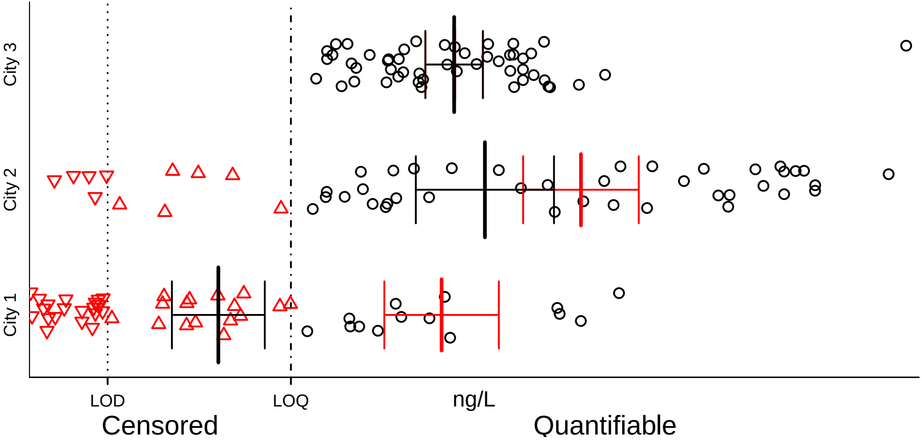
Estimation in the presence of censorship

True mean



Estimation in the presence of censorship

na.rm = T



Estimation in presence of censored data

Less robust solutions

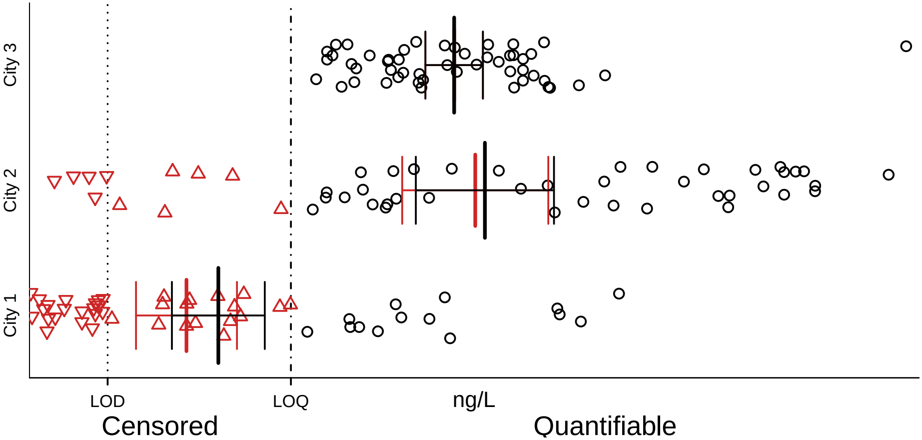
- Dealing with censored data: Substitution

  $<$LOD Below limit of detection $\rightarrow$ plug in 0

  $<$LOQ Below limit of quantification $\rightarrow$ plug in $\frac{1}{2}$ LOQ

Estimation in the presence of censorship

Substitution



Outline

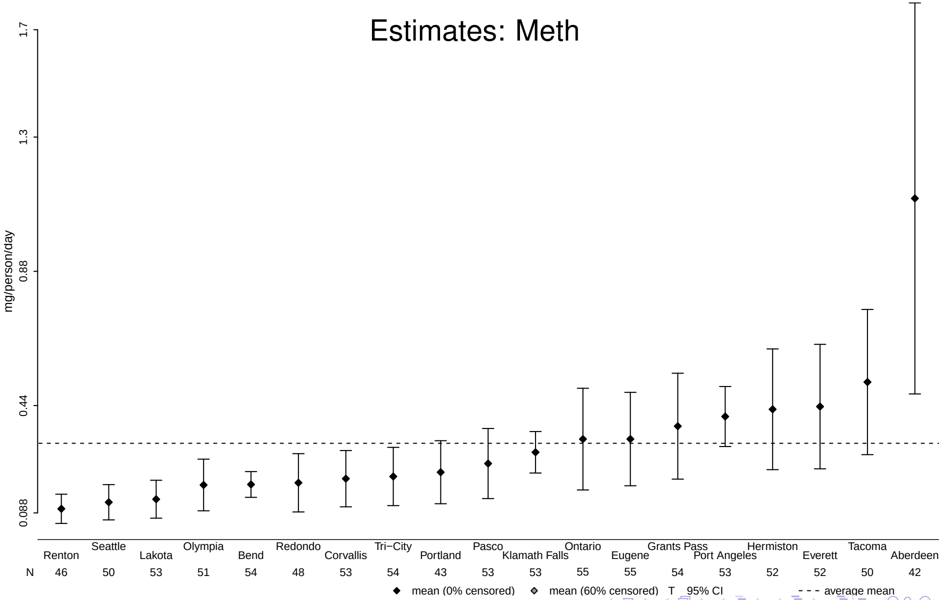
- In *Statistics for censored environmental data using Minitab and R*, Helsel gives guidelines for how to deal with censored observations to create a mean estimate

$< 50\%$	Kaplan-Meier
50–80%	robust Maximum Likelihood Estimation
$> 80\%$	report censoring and perhaps a valid percentile of interest

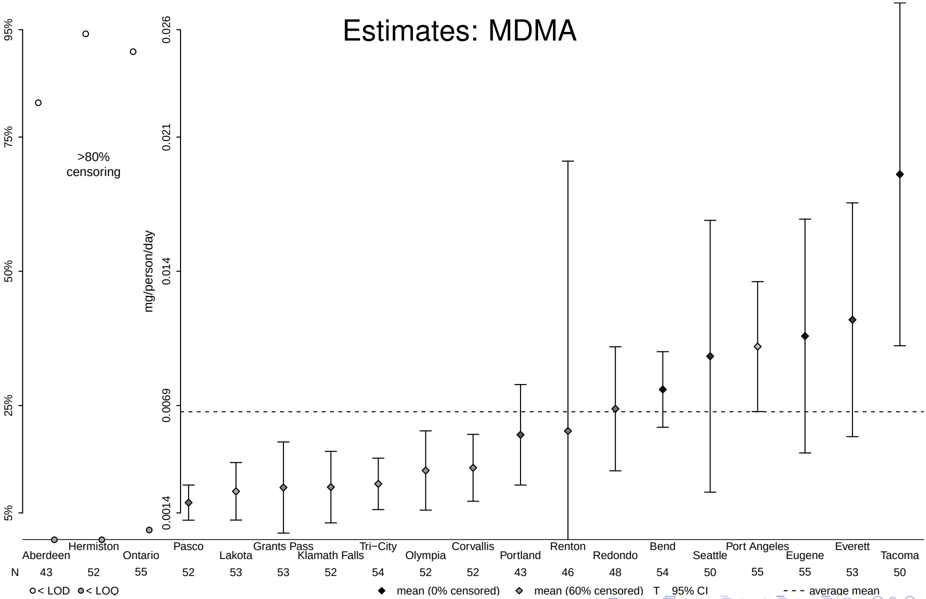
Outline

- Focus on two drugs: methamphetamine and MDMA (Ecstasy)
  - Meth: 1 censored observation (Pasco)
  - MDMA: 3.7% (Bend) to 94.2% (Hermiston), all 19 WWTPs had at least 1 <LOD

Wastewater-based Drug Epidemiology



Wastewater-based Drug Epidemiology



What We Have Added

- We lay out:
  - Building blocks and error components for estimates
  - How to combine them to create estimates with confidence intervals for comparison across place and time
- We improve upon:
  - Population error estimation and accounting for daytime population changes
  - Handling censored data
  - Visualization

Outline

1. Wastewater-based drug epidemiology
   - WBE: Looking upstream to find hidden behaviors
2. From testing results to estimates
   - Assembling the pieces
   - Multiple pieces, multiple uncertainties
   - Estimation in the presence of censored data
   - Apply tools from other fields to generate a sample estimate
3. Visualization and results
   - Examples
4. Summary
   - Your take-aways

How Can You Use This?

- You might improve upon:
  - Error estimation
  - Population estimation for WWTP catchment
    - Mobile phone data? Requires shapefile for catchment.
  - Estimates of metabolism rates and usual doses for given drugs
- Applying censored data methods in other areas with imperfect measurement:
  - HIV viremia counts
  - Population lead levels
  - Other ideas?

How Can You Use This?

- You might improve upon:
  - Population estimation for WWTP catchment
    - Lai, F. Y., et al. (2015). Systematic and Day-to-Day Effects of Chemical-Derived Population Estimates on Wastewater-Based Drug Epidemiology. *Environmental Science & Technology, 49*, 999–1008. `https://doi.org/10.1021/es503474d`
    - O'Brien, et al. (2019). A National Wastewater Monitoring Program for a better understanding of public health: A case study using the Australian Census. *Environment International, 122*, 400–411. `https://doi.org/10.1016/J.ENVINT.2018.12.003`

- Literature suggests using 3 distributions—normal, log, square root—and selecting best fit
- Problem: Retransformation bias is an issue with log transformations, as “the means and variances of the transformed variables are related nonlinearly to the original means and variances, and the process of transforming back gives estimators that often are quite severely biased” (Shumway et al., 2002, p. 3345)

- Problem: Retransformation bias
- Solution: Robust retransformation

Step 1 Take the estimated distribution of the transformed data and place the censored data in appropriate places on the lower part of the distribution. These plotting positions or percentiles are essentially evenly spaced in the lower end of the distribution, on the transformed scale.

  - Specifically, for each censored observation i among all c censored observations within the N observations, the plotting position p is given by

$$p = \frac{c}{N} \times \frac{i - \frac{3}{8}}{c + \frac{1}{4}} \tag{6}$$

    where the i for observations below LOD come before the i for observations above LOD but below LOQ.

- Problem: Retransformation bias
- Solution: Robust retransformation

Step 1 Take the estimated distribution of the transformed data and place the censored data in appropriate places on the lower part of the distribution. These plotting positions or percentiles are essentially evenly spaced in the lower end of the distribution, on the transformed scale.

Step 2 Translate the percentiles into values on the transformed scale via the normal distribution quantile function.

Step 3 Individually re-transform these predicted values on the transformed scale to the original scale.

Step 4 Combine these predicted values with the original observed values (i.e. the uncensored values) into a new set of data.

Step 5 Calculate the mean of this hypothetical data.

- Data exploration with censored data: Say you wanted to do a scatter plot or box plot, what do you do with the censored observations?
  - Could just assign them arbitrary values (e.g., 0 and one-half the LOQ)
  - But avoiding that is why we use censored data methods

- Goal: Data exploration with censored data
- Solution: Use the robust retransformation process to simulate censored data:

Step 1 Take the estimated distribution of the transformed data and place the censored data in appropriate places on the lower part of the distribution. These plotting positions or percentiles are essentially evenly spaced in the lower end of the distribution, on the transformed scale.

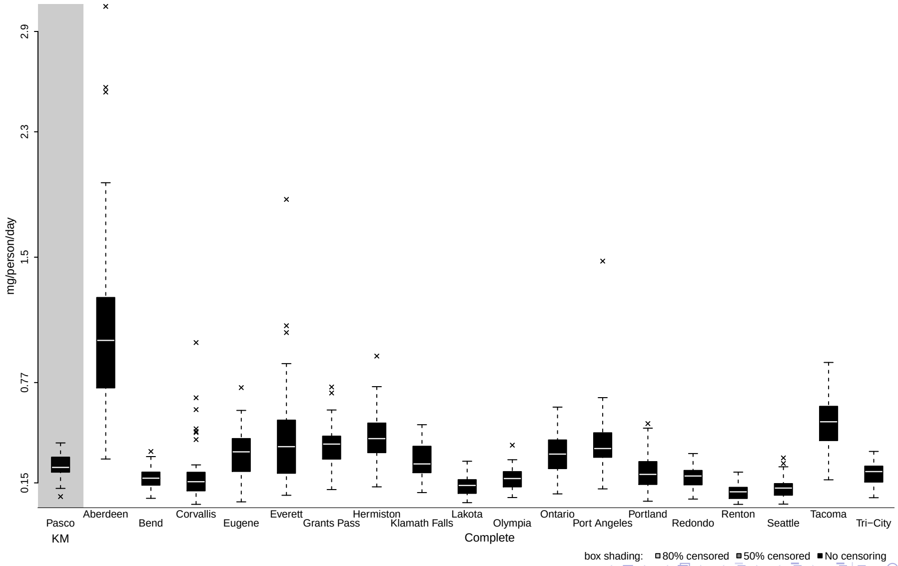
Step 2 Translate the percentiles into values on the transformed (or normal) scale via the normal distribution quantile function.

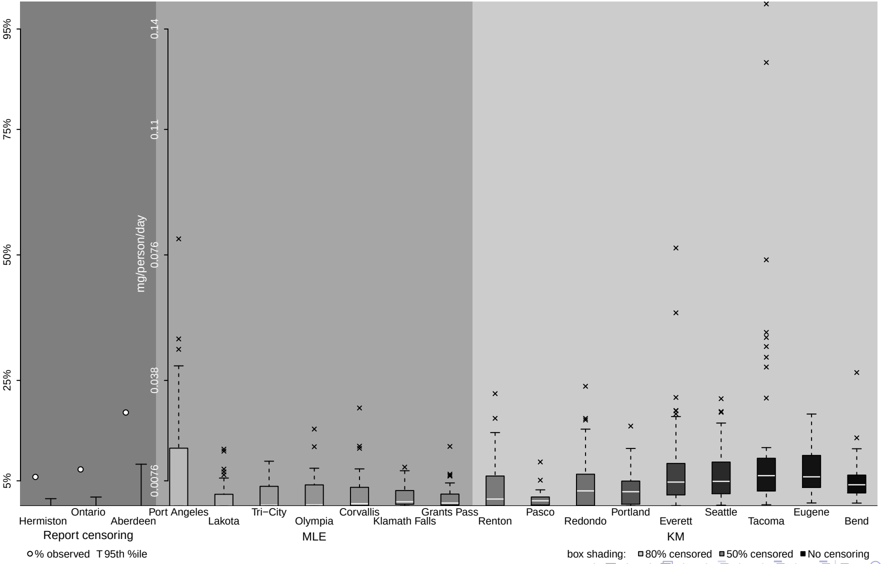
Step 3 Individually re-transform these predicted values on the transformed scale to the original scale (if necessary).

Step 4 Combine these predicted values with the original observed values (i.e. the uncensored values) into a new set of data.

Wastewater-based Drug Epidemiology

Exploration: Meth





95%
75%
50%
25%
5%
0.14
0.11
0.076
0.038
0.0076
mg/person/day
Hermiston
Ontario
Aberdeen
Port Angeles
Lakota
Tri−City
Olympia
Corvallis
Klamath Falls
Grants Pass
Renton
Pasco
Redondo
Portland
Everett
Seattle
Tacoma
Eugene
Bend
Report censoring
MLE
KM
% observed T 95th %ile
box shading: 80% censored 50% censored No censoring