6-20-2016

Using wastewater-based epidemiology to estimate drug consumption—Statistical analyses and data presentation

Caleb J. Banta-Green

Alex J. Brewer

Christoph Ort

Dennis R. Helsel

Jason R. Williams

See next page for additional authors

Authors

Caleb J. Banta-Green, Alex J. Brewer, Christoph Ort, Dennis R. Helsel, Jason R. Williams, and Jennifer A. Field

Using wastewater-based epidemiology to estimate drug consumption—Statistical analyses and data presentation*

Caleb J. Banta-Green[a,*], Alex J. Brewer[b,c], Christoph Ort[d], Dennis R. Helsel[e], Jason R. Williams[a], Jennifer A. Field[b]

[a] *Alcohol and Drug Abuse Institute, University of Washington, Seattle, Washington 98105*
[b] *Department of Environmental Toxicology, Oregon State University, Corvallis, Oregon 97331*
[c] *now Department of Chemistry, Central Washington University, Ellensburg, Washington 98926*
[d] *Eawag, Swiss Federal Institute of Aquatic Science and Technology, CH-8600 Dübendorf, Switzerland*
[e] *Practical Stats, Castle Rock, Colorado 80109*

Abstract

Aim

Analysis of wastewater samples can be used to assess population drug use, but reporting and statistical issues have limited the utility of the approach for epidemiology due to analytical results that are below the limit of quantification or detection. Unobserved or non-quantifiable—censored—data are common and likely to persist as the methodology is applied to more municipalities and a broader array of substances. We demonstrate the use of censored data techniques and account for measurement errors to explore distributions and annual estimates of the daily mean level of drugs excreted per capita.

Measurements

Daily 24-hour composite wastewater samples for 56 days in 2009 were obtained using a random sample stratified by day of week and season for 19 municipalities in the Northwest region of the U.S.

Methods

Methamphetamine, benzoylecgonine (cocaine metabolite), 3,4-methylenedioxymethamphetamine (MDMA), methadone, oxycodone and hydrocodone were identified and quantified in wastewater samples. Four statistical approaches (reporting censoring, maximum likelihood estimation, Kaplan-Meier estimates, or complete data calculations) were used to estimate an annual average, including confidence bounds where appropriate, dependent upon the amount of censoring in the data.

Findings

The proportion of days within a year with censored data varied greatly by drug across the 19 municipalities, with MDMA varying the most (4% to 94% of observations censored). The different statistical approaches each needed to be used given the levels of censoring of measured drug concentrations. Figures incorporating confidence bounds allow visualization of the data that facilitates appropriate comparisons across municipalities.

Conclusions

Results from wastewater sampling that are below detection or quantification limits contain important information and can be incorporated to create a more complete and valid estimate of drug excretion.

*Corresponding author, calebbg@u.washington.edu

*Abbreviations: LOD: limit of detection, LOQ: limit of quantification, KM: Kaplan-Meier estimate, MLE: Maximum Likelihood Estimation, WWTP: wastewater treatment plant, ACS: American Community Survey, RSD: relative standard deviation, MDMA: 3,4-methylenedioxymethamphetamine, CI: confidence interval

Keywords

wastewater; drug epidemiology; wastewater-based epidemiology; censored data; summary statistics; data visualization

1. Introduction

Analyzing wastewater for drugs of abuse and medications dates from the 1970s (Hignite & Azarnoff, 1977) and has increased rapidly since proposed as a source of drug use epidemiology information in 2001 (Daughton, 2001). The first publication of drug consumption estimates was in 2005 (Zuccato et al., 2005). Drug use patterns are becoming more complex as new drugs emerge and drug use expands outside of urban areas. For example, the European Union Early Warning System reported 101 new psychoactive substances in 2014, up from 24 new drugs in 2009 (European Monitoring Centre for Drugs and Drug Addiction, 2015). Expansion of drug use outside of urban areas is exemplified by the recent finding in the United States that fatal drug overdoses have increased 394% in rural areas from 1999–2000 to 2008–2009 (Rossen et al., 2013). To be of value as an epidemiological measure and in turn for policy planning, wastewater drug testing may need to expand to include more compounds in more locations. As wastewater testing expands to cover relatively rare drugs and areas with relatively few users it is possible that an increasing proportion of samples will include results below limit of quantification (LOQ) or detection (LOD), results that are informative and should not be ignored.

While wastewater can be used for binary screening of the presence or absence of a specific urinary biomarker of a substance (e.g., Hernandez et al., 2015; Kinyua et al., 2015), our focus here is on establishing amounts of given substances present. Results are often reported not just in terms of the level of compounds detected, but used as an epidemiological measure of drug use including comparisons of drug trends over time and between locations with implications for international drug policy (Metcalfe et al., 2010; van Nuijs et al., 2009). However, without proper data analysis and reporting of analytical results below LOD and LOQ, as well as estimating confidence bounds that incorporate all sources of uncertainty, we believe reporting such comparisons of drug use are premature. Reporting and analysis issues related to low drug concentrations may become more pronounced as the number of drugs and types of municipalities investigated expand.

A data analysis approach sometimes used is to treat results below LOD as 0 and to replace values below LOQ with a constant; this approach has been shown to lead to biased results (Helsel, 2012), underestimating the mean and either under- or over-estimating the standard deviation. This approach effectively adds a signal to the data that does not truly exist. Appropriate statistical approaches exist for dealing with censored data (data recorded as an interval, or above or below a threshold), which are common in survival analysis (also known as event history analysis). Survival analysis is used in medical studies, for instance to study the impact of a medication on life expectancy when people have incomplete follow-up data, literally an analysis of survival. The same statistical issues exist when data are not available below a certain level (e.g., $x <$LOD), exist between two known points (e.g., LOD$< x <$LOQ), or a dataset contains data with both types of censoring, as is common with environmental samples (Helsel, 2012; Hosmer & Lemeshow, 1999). Helsel provides guidance as to the statistical approach based on both the proportion of the data that are censored and the number of observations and these approaches are implemented in this analysis (2012, p. 92).

To further the development of wastewater testing as a tool for drug epidemiology and informed policy making we implemented a multi-city study with a sampling plan designed to yield an annual estimate of drug excretion. In examining the results of chemical analyses, a substantial proportion of results for some municipalities and some drugs were below LOQ or LOD. Reviewing the existing literature reporting wastewater data yielded no examples of statistical approaches being implemented to address this data issue. However, other areas of environmental chemistry had been exploring these issues for several years (Helsel, 2005). To demonstrate the need for and implementation of these censored data approaches for wastewater-derived drug estimates we present the complete distributions for one year of six compounds across 19 municipalities (results for two substances are detailed in the body of this article, four in the Supplement). In addition, for analyses yielding an annual estimate, we calculate confidence bounds including error components based upon analytical, flow, population and sampling uncertainties. This

data analysis approach yields results that are appropriate for use in epidemiological studies to test for differences in annual drug estimates. Our goal is to demonstrate the steps necessary to convert analytical chemistry results, which may be partially censored, into usable estimates of population drug loads.

2. Methods

When collecting samples from WWTPs we requested that plant operators complete a questionnaire describing the characteristics of the sewer system, 24-hour composite sampling method and the estimated population served. Other data sources, assumptions, and methods are described below.

2.1. Sampling

2.1.1. Locations sampled

A year-long sampling campaign in 19 municipalities in the Northwest region of the United States (the states of Washington and Oregon) was implemented in 2009. Diverse municipalities were chosen for their variable population sizes, commuting patterns, urban or rural location, weather, expected drug use patterns, and resident characteristics in order to maximize variability in the data. Participation by the WWTP was voluntary and no compensation was provided. The participating municipalities were ultimately a convenience sample of willing participants not a representative sample of the entire geographic region. (The locations of the treatment systems, reported population size, and 2009 precipitation are illustrated in Supplement Figure S.1.)

2.1.2. Days sampled to generate annual estimate

To generate a representative annual estimate a time-based, stratified random monitoring approach was utilized that accounted for both seasonality and intra-week (day-to-day) variation. A total of 14 samples each quarter were attempted. At the beginning of the year a schedule was compiled for WWTP staff determining when samples were to be collected: two random Mondays were selected within a quarter, two random Tuesdays etc. This procedure was repeated for each quarter resulting in a total of 56 possible samples over the course of a year covering each day of the week eight times. Based on the drug load variation seen in a few long-term time series for different substances and population sizes this specific monitoring approach has subsequently been shown to result in an annual estimate with an uncertainty not exceeding approximately 10% (Ort et al., 2014a; see also Supplement Figure S.7). Monitoring days were the same for all 19 locations. Of the attempted 56 samples for each of the 19 locations, between 42 and 55 were actually collected, resulting in a total of 971 (91% of possible 1064).

2.1.3. Daily composite sampling

Each WWTP provided 24-hour composite samples. The approach to compositing varied by municipal WWTP in the time intervals, averages, and minimum and maximum sampling frequencies, as well as whether flow, time, or volume proportional sampling was utilized. The time of day at which sampling began and ended varied across plants.

2.2. Data reporting and sources

Total wastewater volumes entering the plant were documented by WWTPs. Analytical methods, including internal standards and stability of analytes in acidified samples, are detailed in Chiaia and colleagues (2008). Briefly, wastewater samples were acidified (with HCl) and shipped for analysis. Samples were titrated to neutrality (with NaOH) and a titration factor (final volume ÷ initial volume) was recorded. Large volume injection liquid chromatography/tandem mass spectrometry was used to identify and quantify illicit drugs and metabolites. Compounds studied included methamphetamine, 3,4-methylenedioxymethamphetamine (i.e. MDMA, Ecstasy), and benzoylecgonine, a major metabolite of cocaine, substances we knew to be detectable and to show high variability across the region (Banta-Green et al., 2009). In addition, the opioid medications methadone, hydrocodone and oxycodone were studied as they are widely prescribed and used licitly and illicitly. Both methamphetamine and cocaine have very low levels of pharmaceutical use in the U.S., for appetite suppression and anaesthesia respectively, with the vast majority of consumption being illicit. MDMA is an illicit drug. The opioids are commonly used for pain control and, in the case of methadone, addiction treatment. At the time of the study the

illicit availability and use of these opioids was relatively high (Dart et al., 2015). The limits of detection and quantification were established with respect to a signal to noise ratio of ≥ 3 and ≥ 10, respectively, as described in Chiaia et al. (2008), and are detailed in Supplement Table S.1. For illustration purposes methamphetamine and MDMA are the focus in this paper, with figures for the other substances provided in Supplement Figures S.2 to S.5.

3. Calculation: Analysis in the presence of censoring and uncertainty

Our goal here is to demonstrate the full use of the information available to provide useful comparative data across place and time. We estimate the average index or per capita load of analytes of interest for one year, based on 56 possible samples stratified through the year, for a given municipality-drug combination, that is a specific substance for a given WWTP catchment area. Others have reported per capita estimates of consumption of drugs and methods for such calculations have been described (e.g., Zuccato et al., 2005). However, we are reporting estimates of per capita excretion of drugs as fewer assumptions are needed, e.g., purity of parent compounds, typical doses, routes of administration, and metabolization rates of drugs *in vivo* (see, e.g., Bruno et al., 2014).

The estimated index load represents the average of the individual daily observations:

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

Each individual observation i of the N observations for a municipality-drug incorporates several parts, represented by the equation below (which leaves out scaling constants to produce a result in the desired units):

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

- C_i is the estimated concentration in the sample, possibly with censoring (indicated by unique codes to indicate <LOD or <LOQ)
- T_i is a titration factor indicating how the sample was modified to facilitate chemical analysis, assumed to be measured without error
- F_i is the estimated flow of liquid through the WWTP for that day
- P_i is the estimated population of users of the WWTP for that day

The sampling and testing procedure produced slightly less than 56 samples per location, as individual samples were unavailable for various reasons (e.g., a test tube broke during shipping or a sample was not submitted). We treat these as missing completely at random. Where the remaining observations are complete, i.e. all quantified (≥LOQ), the average in equation 1 can be calculated directly.

With incomplete or censored information, we follow the guidelines of Helsel (2012) and use where possible methods based on survival analysis, analyzing loads as though they represented times with left-censoring (<LOD) and interval censoring (between LOD and LOQ). These methods require simply transforming observed quantities and LOD and LOQ, as relevant, into intervals representing the lowest and highest possible value for that observation to represent what we do know about that observation (sample data setup is available in Supplement section S 9.3). Specifically, for moderate censoring (<50% of the samples), we use the non-parametric Kaplan-Meier (KM) estimator to produce the mean, via the `Surv` function of the R package `survival`.

For medium censoring (50% to 80% of the samples), we add assumptions about the underlying distribution of the real data and use parametric Maximum Likelihood Estimation for censored data. Following advice in the literature on analysis of censored environmental data (Helsel, 2012; Shumway, Azari, & Kayhanian, 2002), we restrict our distributional choices to the log-normal (transforming via the natural log and assuming a normal distribution for the transformed observations), the square root (creating transformed values $y_i = 2(\sqrt{x_i} - 1)$ and assuming a normal distribution for the transformed y_i), and normal (no transformation), and select the result that best fits the data (highest log-likelihood). We produce

MLE estimates using the `fitdistcens` function of the R package `fitdistrplus`, and then re-transform the result via a robust retransformation to avoid retransformation bias (Helsel, 2012; Kroll & Stedinger, 1996; Shumway, Azari, & Kayhanian, 2002). 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), although we apply the method in cases where the square root transformation was chosen as well. This method (described further in Supplement section S 9.4) involves randomly placing the censored observations in the appropriate range of the probability distribution, converting each individual observation from a probability to a number on the transformed distribution, reversing the transformation individually for each of the censored observations and merging them with the observed values on the original scale, and then calculating the mean and variance.

For drugs and municipalities with greater than 80% of samples censored, simply reporting the proportion of samples with quantifiable results is appropriate. One may also report percentiles above the proportion censored: For example, with 94% censoring (as is the case with MDMA in Hermiston in the Results below), one should report the proportion below LOD and below LOQ and perhaps the 95^{th} or 99^{th} percentiles.

3.1. Population estimates

As noted above, equation 2 divides the estimated total amount of the analyte by an estimate of the number of users of the WWTP for that day. For regulatory and planning purposes a WWTP needs to have an estimate of the population served by the treatment plant. The questionnaire asked the basis for the population estimate and common responses included the US Census Bureau, the number of sewer connections adjusted by a multiplier, or official state estimates (themselves based on Census data). These are estimates of a typical population over the course of a year, which may not correspond with the year sampled, and may match the actual number of users with unknown accuracy. For example, Census-based estimates are based on political boundaries, such as city boundaries, whereas WWTP catchment areas rarely align perfectly with political boundaries and will often take advantage of topographical features to maximize gravity-based flows. Survey responses exhibited various levels of precision (e.g., 10,000 versus 15,980).

Where possible, we account for differences between daytime population and baseline residential population by applying American Community Survey 2006–2010 estimates of daytime population (which are calculated only for those who worked in the week prior to the ACS and only in regards to going to and from work, thus ignoring travel for any other reason). Specifically, we apply the ACS estimate of proportional population change to the reported population size to estimate the net daytime in- or out-migration and add half of this amount to the reported population size, to represent work hours being approximately half of waking (and thus waste-producing) hours during the day. We are assuming that the proportional change from ACS is geographically representative and thus can be applied uniformly even when the WWTP may serve a population slightly smaller than the resident population of the same-named town. This adjustment is not possible for WWTPs that do not correspond reasonably to Census-recognized places, as with Tri-City (three Portland suburbs plus part of a fourth), Lakota (a neighborhood within the city of Federal Way, WA), Redondo (a neighborhood partly in Federal Way and partly in Des Moines, WA), Seattle (the WWTP of which serves several suburbs), and Renton (a WWTP that serves several cities and unincorporated areas). In addition, where the survey responses indicated the source of the population estimate was "connections", as in Grants Pass, Olympia, and Tri-City, we assumed this did indeed account for users of both residential and commercial toilets.

An additional complexity exists for population estimation for the Seattle and Renton WWTPs. Unlike others in this sample of WWTPs, the catchment area of these systems varies depending upon loads and rainfall. Specifically, during certain periods, a switch sends flow from certain pumping stations to the Renton WWTP instead of the Seattle WWTP. Our daily population estimate for sampled days for these two systems reflects the reported position of this switch on those days. For all other cities, P_i in equation 2 is constant. (Supplement Table S.2 reports the source for and the final estimate of population size. The populations for Seattle and Renton reflect the average across all sampled days.)

3.2. Uncertainty analysis

Our final estimates are composed of four elements, three of which are directly measured with error and one of which is an indirect estimate with inherent uncertainty. Each of the components of an individual observation in equation 2 could conceivably be measured with error, although we follow standard practice and assume titration is error-free. For the remaining components, future efforts may improve measurement and estimation of the associated measurement error. In addition, in using a sample of days to represent 365 days, there is error associated with creating an annual sample. Uncertainty associated with a composite sample within a day representing the whole day is subsumed in estimates of within-year sampling error. We thus have four sources of uncertainty, associated with sampling, chemical analysis, flow measurement, and population estimation, which are commonly assumed to be independent (e.g., Lai et al., 2015; Ort et al., 2014b, 2010a) as no physical or conceptual dependence has been reported.

We follow prior work (e.g., Brewer et al., 2012; Lai et al., 2011) for this demonstration and combine the error components by

$$U_T = \sqrt{U_S^2 + U_C^2 + U_F^2 + U_P^2} \tag{3}$$

where each uncertainty component, and the total uncertainty, is expressed as a coefficient of variation or relative standard deviation (RSD). The total uncertainty is necessary for estimating confidence intervals and interpreting differences between municipalities. Multiplying U_T by the estimated yearly mean produces a standard deviation, which can be combined with the t-distribution value appropriate for the number of samples and the desired confidence level to produce a confidence interval.

3.2.1. Sampling error

Researchers typically rely on WWTPs' existing 24-hour composite sampling for the sub-samples obtained for analysis due simply to the dramatic savings in resources compared with researchers directly collecting their own samples. WWTPs with which we and others have worked use a wide range of sampling approaches in terms of compositing approach, e.g., flow-, time- or volume-weighted, as well as the frequency of sampling. Few plants with which we have worked use ideal sampling approaches and few are willing to substantially change their sampling protocols for the extended sampling campaigns we have conducted for various research projects. As described by Ort and colleagues (2010b), substantial error can result based on how within-day sampling is carried out and the frequency with which it is conducted.

Our annual sampling frame was informed by the types of variability drug users experience in their lives based on biopsychosocial and financial circumstances. Ort and colleagues (2014a) estimate an expected uncertainty due to sampling (or monitoring) of approximately 10% with approximately 56 samples stratified by quarter of year and weekday versus weekend. (Sensitivity analyses using half and twice this error for MDMA appear in Supplement section S 9.6; similar graphs for the full slate of analytes are available upon request.)

3.2.2. Concentration error

The reported concentration of the substance in the sample is inherently measured with error, which is partially represented by the detection and quantification limits themselves. Above these levels, there remains uncertainty in the analytical method. For concentration we use the RSDs from the testing procedures for each municipality-drug combination, which resulted from testing the same sample 2 to 4 times (Brewer et al., 2012; Chiaia et al., 2008).

3.2.3. Flow error

The daily flow of wastewater to a plant is incorporated in order to account for variability and dilution due to system usage due to human and weather factors. While WWTPs measure flow and generally have an estimate of the error associated with the flow, often manufacturers' specifications, it is believed these estimates are too small. We use an estimate of 5% from prior WWTP testing protocols (Brewer et al., 2012). Investigating the true nature of the error associated with the flow estimate is worthwhile and should be utilized whenever possible in real world applications.

3.2.4. Population uncertainty

In addition to the daytime population growth described above, population estimates retain myriad sources of further error. Sampling-based methods (i.e. the ACS) include error estimates in their results. For the 4 WWTPs that reasonably correspond to Census-defined areas and which provided Census-based estimates (Aberdeen, Bend, Klamath Falls, and Port Angeles), we used 5-year ACS error estimates for the underlying population. These 4 WWTPs also had the daytime population correction applied, itself with an estimated error, to produce a combined population error estimate (ranging from 2.9% to 10.4%).

For other possible survey responses, we have no way to estimate any error associated with the initial population estimate. For example, estimating the accuracy of the multipliers WWTPs use for planning purposes is beyond the scope of the current article. Corvallis, Hermiston, Portland, and Pasco reported that their population estimates came from official state estimates, which do not provide error estimates. We did not feel comfortable assuming 0 measurement error for resident population given their estimation procedures. As a result, 15 of our WWTP systems have associated population error estimates of 20% as in Ort and colleagues (2014b). ACS 2006–2010 population estimates for the 14 WWTPs that reasonably correspond to Census areas have RSDs ranging from 0.68% to 20.08%, suggesting 20% is a reasonably conservative estimate here.

4. Results

4.1. Censoring and distributions

Censoring ranged from 0 to 94.2% across the $(19 \times 6 = 114)$ municipality-drug combinations. Methamphetamine exhibited the least censoring, with only Pasco having a single unobservable observation. MDMA exhibited the most censoring, with no WWTP having complete observations. Censoring of MDMA observations ranged from 3.7% (Bend) to 94.2% (Hermiston), and every municipality had at least one observation below LOD. These patterns of missingness are likely related to typical usage patterns, where we might expect MDMA to be present at low quantities, if at all, on weekdays, but expect methamphetamine use to be more uniform across days of the week (Andrés-Costa, Rubio-López, Morales Suárez-Varela, & Pico, 2014; Banta-Green et al., 2005; Simon et al., 2002).

In order to present the distribution of observations in the presence of censoring, we utilize the hypothetical data created by the robust retransformation process described above (and further in Supplement section S 9.4), applied even if the amount of censoring indicates use of the KM estimate, in order to create an approximation of what the distribution might have been in the absence of censoring. In Figure 1 we present standard boxplots of these observations, with WWTPs arranged in order from most to least censoring (from left to right; alphabetically among those with the same amount of censoring), for methamphetamine (top panel) and MDMA (bottom). The amount of censoring is further indicated by the shading of the box, ranging from lightest gray to black, the latter indicating complete data. Labels and background shading indicate which statistical method is utilized (Report censoring if $> 80\%$ censoring, MLE if 50–80% censoring, KM if $< 50\%$ censoring, Complete data if no censoring) to create the estimated annual mean in the next section. (Precise levels of censoring are available in Supplement Table S.2.) Where there is no background shading and all the boxes are black, the observed complete data is used to create the boxplot (and the estimated mean). Where there is censoring, as for Pasco methamphetamine (left side of top panel) and MDMA in all cities (bottom panel), the boxplots reflect a mix of observed data and data created via the robust retransformation.

The re-created data is obviously at or below the lower end of the box. As the points are placed somewhat arbitrarily on the low end of the distribution, the actual position of the lower whisker (e.g., Bend MDMA in Figure 1), or of the first quartile where there is 25% or more censoring (e.g., Portland MDMA, with 32.6% censoring), should not be interpreted strongly in the presence of more than minimal censoring. As per the guidelines above, we do not present boxplots for the three municipality-drug combinations, MDMA in Aberdeen, Hermiston, and Ontario, in which there was over 80% censoring. Instead, we report for these cases the percentage observed (left-most axis; the complement of the percentage censored) and the 95^{th} percentile, as Hermiston had just less than 95% censoring.

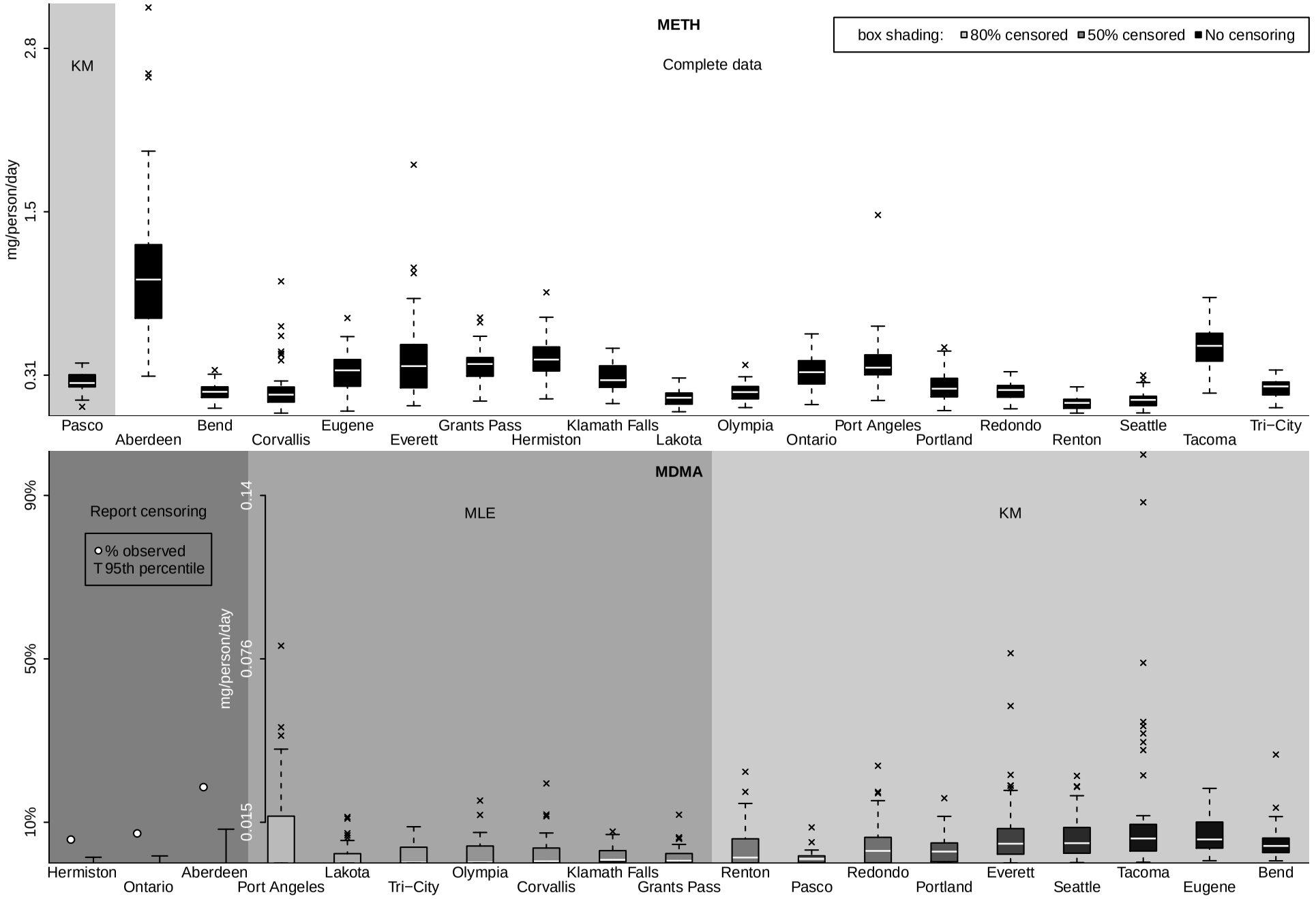

Figure 1: Methamphetamine (top panel) and MDMA (bottom) index load distribution (mg/person/day)

The boxplots present the median of the (possibly re-created) data, as a white line across the box, and one definition of outliers, marked as $\times$. Specifically, each whisker extends to the most extreme data point no more than 1.5 times the interquartile range—the length of the box—and any point beyond the whisker is an outlier. For municipality-drug combinations with more than 50% but less than 80% censoring, for which the robust MLE is used to create the estimate below, this median is highly dependent on the re-created data from the robust retransformation. Consistent with the amount of censoring for such cases, the medians tend to be near 0 (and, given scaling, may be difficult to differentiate from 0 in the figures, as for the first four box plots—Port Angeles through Olympia—in the bottom panel of Figure 1). The figures are scaled so that the highest observation across the 19 cities defines the top of the range, with the tick marks representing 10%, 50%, and 90% of this range. We see for Port Angeles MDMA, for example, that the nearly 80% censoring is indicative of many days with very low, possibly 0, levels of MDMA, but that Port Angeles also had one outlier measured at over half of the highest observation among all WWTPs (for Tacoma).

4.2. Annual estimates with error bounds

While the boxplots represent the *distributions* of observations across the year, or hypothetical distribution in the presence of censoring, in this section we focus on the final *estimates* of annual average daily index load. In addition, we focus on the error bounds created by the above combination of the error components, to give an estimate of the uncertainty around the mean. Figure 2 orders the WWTPs from lowest to highest estimate, where the *estimated mean* is represented by a diamond shaded according to the amount of censoring (as with the boxplots in the previous section). The 95% confidence interval (CI) is represented by the T-shaped line extending above and below each diamond. Below each WWTP name we present the number of samples used to create the estimate and the CI, including those observations included in the estimate despite censoring. The dotted line represents the (unweighted) average of the annual estimates of the 19 WWTPs.

For methamphetamine (Figure 2, top panel), the average index load ranges from 0.10 milligrams per person per day in Renton to 1.12 mg in Aberdeen, an 11-fold difference. The 95% CIs allow us to say that, within the bounds of our error estimation, Aberdeen has significantly more methamphetamine in its wastewater than Renton, and significantly more than the average WWTP. Renton, Seattle, Lakota, Olympia, Bend, Redondo, Corvallis, and Tri-City all have below average levels.

For MDMA, again, we present only information on censoring for the three highly censored WWTPs, on the left in the bottom panel of Figure 2. The municipalities with the lowest seven estimates are all significantly below average, while Tacoma and Port Angeles have above average annual mean MDMA levels. Given the error combination approach used here (Equation 3) and that the other error components are small and/or constant ($\leq$ 20%), the large confidence intervals in Figure 2 are due primarily to analytical error associated with chemical testing—for example, the analytical error for MDMA for Renton had an RSD of 121%, while Grants Pass MDMA testing had an RSD of 37%.

5. Discussion

Municipal sewage-based testing for medications and drugs of abuse is moving into its adolescence, but is not fully mature in part due to the lack of an adequate statistical treatment of the data resulting from chemical analysis. The ability to document the presence of substances has been shown repeatedly around the world (e.g., Banta-Green et al., 2009; Castiglioni et al., 2006; Zuccato et al., 2005). However, the utility and interpretability of these data is diminished until summary measures, such as means, accurately and completely reflect the collected data including data below limits of quantification and/or detection. Current efforts to interpret results are further weakened by the inconsistent use of confidence bounds around estimates and missing components for estimation of uncertainty, e.g., related to population size or the limited number of observations. Without these comprehensive confidence bounds it is not possible to indicate whether a measured level of a substance (or its metabolite) is higher than a prior estimate for the same substance or different compared to another municipality. In other words, current presentations of data have limited utility as epidemiological data and in turn are of limited value for research, practice or policy purposes.

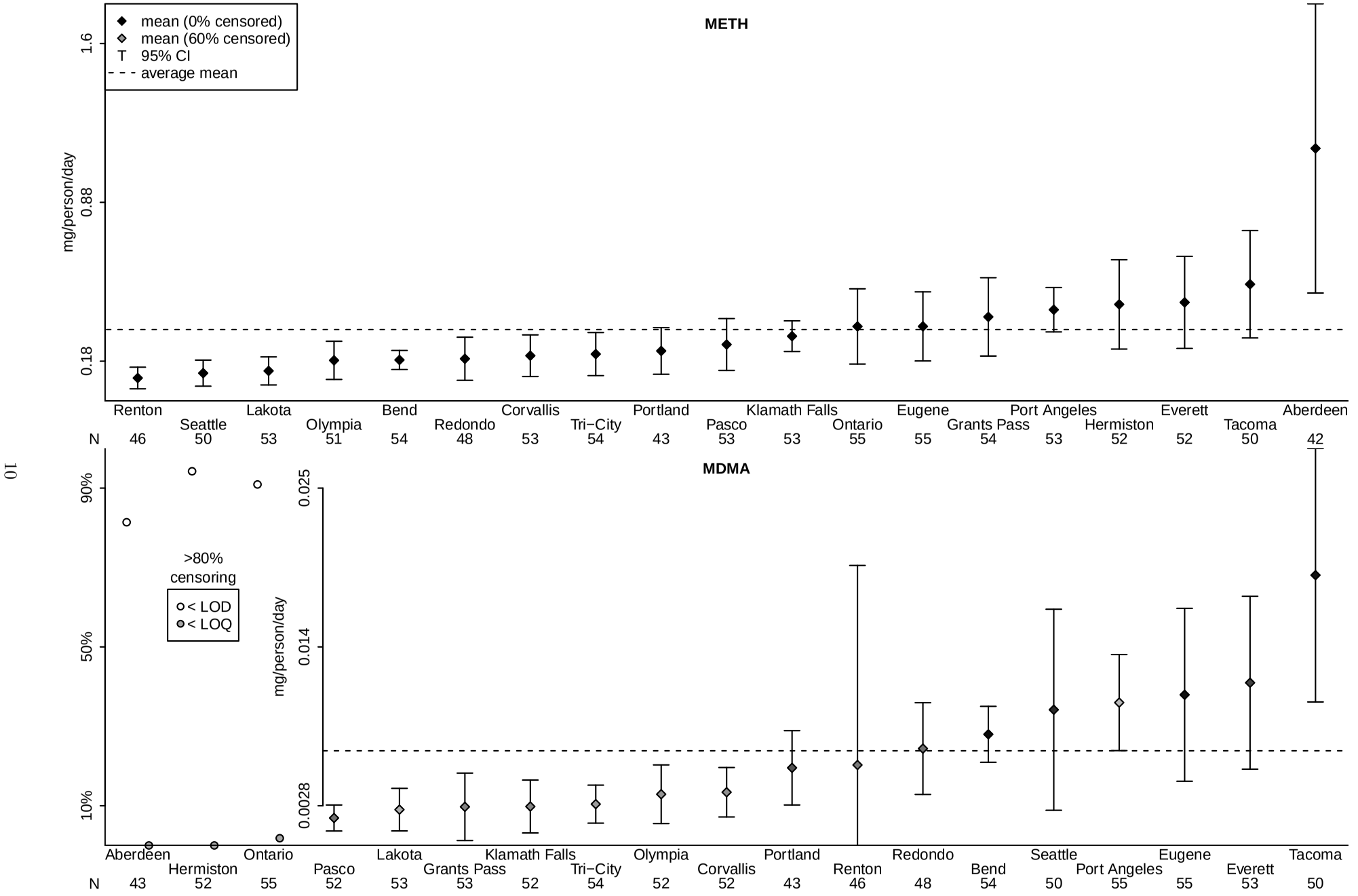

Figure 2: Methamphetamine (top panel) and MDMA (bottom) index load mean with 95% confidence interval (mg/person/day)

Our goal in this paper was to illustrate a more complete approach to analyzing wastewater data. Rather than presenting annual estimates as though all the underlying components were measured without error, we have incorporated uncertainty estimation and have been transparent about sources of error and when we had to make assumptions about the size of that error. Rather than substituting an arbitrary value, we used statistical methods that use the information inherent in knowing a value is between 0 and the LOD or between the LOD and LOQ to generate an unbiased estimate of the mean annual per capita load.

We believe that an approach such as that presented here provides more accurate and valid results and encourage its use, and refinement, in future analyses. Previously published data which did not incorporate these statistical approaches may have average values that are incorrect and biased. The actual computed estimate may, or may not, change significantly with our approach, but it will be less biased and more valid. Further, previously published analyses that reported significant differences between places or over time may have overstated differences due to incomplete, and incorrectly small, confidence intervals. Assuming the necessary data elements are available, these data could be reanalyzed with the approaches suggested here and yield data valuable in longitudinal comparisons.

We have also demonstrated the difficulty surrounding the seemingly simple designation of a denominator for estimating per capita (or per 100,000) loads. Some have preferred the term "index load" in lieu of "per capita load" to capture the uncertainty around estimation of the denominator (Chiaia et al., 2008). We show how available survey data on daytime population growth due to commuting could be applied to adjust a given population estimate. Future work should investigate improving upon the underlying population estimation. A revised and more complex WWTP operator survey might, for example, request that WWTP operators be more specific in their estimate of population size and the source of that estimate. The researcher may need to facilitate a conversation between the WWTP staff and city planners and engineers to understand system and population characteristics in more detail. WWTP coverage areas could be approximately matched with census blocks or tracts to create better estimates with better measurement error approximations, although daytime population change estimates may be available only for larger towns and cities. Work is ongoing to try to better estimate the size of the underlying population using a range of other measureable analytes (see, e.g., Brewer et al., 2012; Castiglioni et al., 2013; Lai et al., 2015, 2011; Ort et al., 2014a). Finally, researchers could obtain data from mobile phone companies during the average day and average night to understand daily migration patterns.

As noted throughout, further refinement of measurement and error estimation of not only population but all parameters will help mature the use of WWTP sampling for research and policy regarding illicit and licit drug and medication use as well as other target analytes yet to be determined. For example, our analytical errors were relatively small for commonly used drugs such as methamphetamine (where analytical RSDs ranged from 0.4% to 23.9%, with an unweighted mean of 4.8%), but much larger for relatively rare MDMA (range 1.2% to 121.1%, unweighted mean 19.4%). If the sample selected for repeated testing was from a Sunday, more MDMA would be expected to be present (due to Saturday consumption) and thus more consistently quantifiable than with a Tuesday sample. Similarly, the sampling error estimate of 10% (Ort et al., 2014a) was derived from repeated samples of $N = 56$ taken from over 1000 consecutive daily samples of benzoylecgonine (a metabolite of cocaine) in a small WWTP serving approximately 7000 inhabitants. Measuring a substance with higher seasonal variation would presumably increase the error estimate associated with 56 samples. Substituting a larger or smaller error component estimate has relatively small effects on the resulting confidence intervals in comparison to inter-municipality variation (sensitivity analyses using sampling error of 5% and 20% for MDMA appear in Supplement section S 9.6; others available upon request). Furthermore, more advanced methods of estimating error via Monte Carlo simulation are available, but the linear error propagation used here (equation 3) is generally thought to be an acceptable approximation especially if individual error components are not too large (Lai et al., 2015; Ort et al., 2014b). Finally, the results presented here are index loads of excreted substances, not estimates of the number of users or the average ingestion of a given parent substance. Presenting such results relies on further assumptions about factors such as excretion, absorption, and dosage, and their relevant measurement errors.

A limitation of the current analysis, and one that may impact others relying upon WWTP operators to provide samples, is that no location provided all 56 samples. Between 42 and 55 samples were provided

by each location. Overall 91% of possible samples were submitted, potentially lessening statistical power.

6. Conclusion

The ability to utilize wastewater-derived data for practice and policy decision making is important as drug use continues to have major public health consequences (Jenkins et al., 2011; Jones et al., 2015, 2013), and the methodology could be readily applied to other target analytes that are anthropogenic markers of difficult to measure human behaviors. Existing methods for drug surveillance are crude (National Institute on Drug Abuse, 2006). Drug abuse epidemiology currently relies on indicators that are limited by: 1) poor geographic detail, 2) an emphasis on major metropolitan areas 3) significant time lag in data availability, 4) detection and self-report biases, 5) measurement error, and 6) an over-reliance on morbidity and mortality data.

The need for a complete statistical approach for wastewater samples was demonstrated by the highly variable distributions of quantifiable results for different compounds and across locations. A comprehensive approach to developing an estimate that includes information contained in results below quantification and detection as well as error bounds is presented. The results can be used to test for differences in drug levels across places. Although the resulting estimates are of index loads of excreted substances, we believe examining such results across place and/or time is a highly useful addition to other drug surveillance sources given the dearth of local quantifiable data. This general statistical approach can be used in the future even as methods for determining the various error components improve.

7. Acknowledgments

The authors thank the wastewater treatment plant operators who provided data for this study. Funding for data collection was provided by the National Institute on Drug Abuse to Drs. Banta-Green and Field (R21 DA024800); the findings and opinions expressed in this paper are those of the authors.

8. References

Andrés-Costa, M.J., Rubio-López, N., Morales Suárez-Varela, M., Pico, Y., 2014. Occurrence and removal of drugs of abuse in Wastewater Treatment Plants of Valencia (Spain). Environ. Pollut. 194, 152–162. http://dx.doi.org/10.1016/j.envpol.2014.07.019

Banta-Green, C., Goldbaum, G., Kingston, S., Golden, M., Harruff, R., Logan, B.K., 2005. Epidemiology of MDMA and associated club drugs in the Seattle area. Subst. Use Misuse 40, 1295–1315.

Banta-Green, C.J., Field, J.A., Chiaia, A.C., Sudakin, D.L., Power, L., De Montigny, L., 2009. The spatial epidemiology of cocaine, methamphetamine and 3,4-methylenedioxymethamphetamine (MDMA) use: A demonstration using a population measure of community drug load derived from municipal wastewater. Addiction 104, 1874–1880.

Brewer, A.J., Ort, C., Banta-Green, C.J., Berset, J.-D., Field, J.A., 2012. Normalized Diurnal and Between-Day Trends in Illicit and Legal Drug Loads that Account for Changes in Population. Environ. Sci. Technol. 46, 8305–8314.

Bruno, R., Hall, W., Kirkbride, K.P., Lai, F.Y., O'Brien, J.W., Prichard, J., Thai, P.K., Mueller, J.F., 2014. Commentary on Ort et al. (2014): What next to deliver on the promise of large scale sewage-based drug epidemiology? Addiction 109, 1353–1354. http://dx.doi.org/10.1111/add.12651

Castiglioni, S., Bijlsma, L., Covaci, A., Emke, E., Hernández, F., Reid, M., Ort, C., Thomas, K. V, van Nuijs, A.L.N., de Voogt, P., Zuccato, E., 2013. Evaluation of Uncertainties Associated with the Determination of Community Drug Use through the Measurement of Sewage Drug Biomarkers. Environ. Sci. Technol. 47, 1452–1460. http://dx.doi.org/10.1021/es302722f

Castiglioni, S., Zuccato, E., Crisci, E., Chiabrando, C., Fanelli, R., Bagnati, R., 2006. Identification and Measurement of Illicit Drugs and Their Metabolites in Urban Wastewater by Liquid Chromatography–Tandem Mass Spectrometry. Anal. Chem. 78, 8421–8429. http://dx.doi.org/10.1021/ac061095b

Chiaia, A.C., Banta-Green, C., Field, J., 2008. Eliminating Solid Phase Extraction with Large-Volume Injection LC/MS/MS: Analysis of Illicit and Legal Drugs and Human Urine Indicators in US Wastewaters. Environ. Sci. Technol. 42, 8841–8848. http://dx.doi.org/10.1021/es802309v

Daughton, C.G., 2001. Illicit drugs in municipal sewage. Proposed new nonintrusive tool to heighten public awareness of societal use of illicit-abused drugs and their potential for ecological consequences. ACS Symp. Ser. 791, 348–364.

European Monitoring Centre for Drugs and Drug Addiction, 2015. New psychoactive substances in Europe: An update from the EU Early Warning System. http://dx.doi.org/10.2810/372415

European Monitoring Centre for Drugs and Drug Addiction, 2016. Assessing illicit drugs in wastewater: Advances in wastewater-based drug epidemiology (Insights 22). Publications Office of the European Union, Luxembourg. http://dx.doi.org/10.2810/6622

Helsel, D.R., 2012. Statistics for censored environmental data using Minitab and R, 2nd ed. Wiley, Hoboken, N.J.

Helsel, D.R., 2005. More than obvious: Better methods for interpreting nondetect data. Environ. Sci. Technol. 39, 419A–423A. http://dx.doi.org/10.1021/es053368a

Hernandez, F., Ibañez, M., Botero-Coy, A.-M., Bade, R., Cristina Bustos-Lopez, M., Rincón, J., Moncayo, A., Bijlsma, L., 2015. LC-QTOF MS screening of more than 1,000 licit and illicit drugs and their metabolites in wastewater and surface waters from the area of Bogota, Colombia. Anal. Bioanal. Chem. 407, 6405–6416. http://dx.doi.org/10.1007/s00216-015-8796-x

Hignite, C., Azarnoff, D.L., 1977. Drugs and drug metabolites as environmental contaminants: Chlorophenoxyisobutyrate and salicylic acid in sewage water effluent. Life Sci. 20, 337–341.

Hosmer, D.W., Lemeshow, S., 1999. Applied Survival Analysis: Regression Modeling of Time to Event Data, 1st ed. John Wiley & Sons, Inc., New York.

Jenkins, L.M., Banta-Green, C.J., Maynard, C., Kingston, S., Hanrahan, M., Merrill, J.O., Coffin, P.O., 2011. Risk factors for nonfatal overdose at Seattle-area syringe exchanges. J. Urban Heal. 88, 118–128.

Jones, C.M., Logan, J., Gladden, R.M., Bohm, M.K., 2015. Vital Signs: Demographic and Substance Use Trends Among Heroin Users—United States, 2002–2013. MMWR. Morb. Mortal. Wkly. Rep. 64, 719–725.

Jones, C.M., Mack, K.A., Paulozzi, L.J., 2013. Pharmaceutical overdose deaths, United States, 2010. JAMA 309, 657–659. http://dx.doi.org/10.1001/jama.2013.272

Kinyua, J., Covaci, A., Maho, W., McCall, A.-K., Neels, H., van Nuijs, A.L.N., 2015. Sewage-based epidemiology in monitoring the use of new psychoactive substances: Validation and application of an analytical method using LC-MS/MS. Drug Test. Anal. 7, 812–818. http://dx.doi.org/10.1002/dta.1777

Lai, F.Y., Anuj, S., Bruno, R., Carter, S., Gartner, C., Hall, W., Kirkbride, K.P., Mueller, J.F., O'Brien, J.W., Prichard, J., Thai, P.K., Ort, C., 2015. Systematic and Day-to-Day Effects of Chemical-Derived Population Estimates on Wastewater-Based Drug Epidemiology. Environ. Sci. Technol. 49, 999–1008. http://dx.doi.org/10.1021/es503474d

Lai, F.Y., Ort, C., Gartner, C., Carter, S., Prichard, J., Kirkbride, P., Bruno, R., Hall, W., Eaglesham, G., Mueller, J.F., 2011. Refining the estimation of illicit drug consumptions from wastewater analysis: Co-analysis of prescription pharmaceuticals and uncertainty assessment. Water Res. 45, 4437–4448. http://dx.doi.org/10.1016/j.watres.2011.05.042

Metcalfe, C., Tindale, K., Li, H., Rodayan, A., Yargeau, V., 2010. Illicit drugs in Canadian municipal wastewater and estimates of community drug use. Environ. Pollut. 158, 3179–3185.

National Institute on Drug Abuse, 2006. Epidemiologic Trends in Drug Abuse: Highlights and Executive Summary, in: Proceedings of the Community Epidemiology Work Group, Volume I. Division of Epidemiology, Services and Prevention Research, National Institute on Drug Abuse, Bethesda MD, pp. 1–80.

Ort, C., Eppler, J.M., Scheidegger, A., Rieckermann, J., Kinzig, M., Sörgel, F., 2014a. Challenges of surveying wastewater drug loads of small populations and generalizable aspects on optimizing monitoring design. Addiction 109, 472–481. http://dx.doi.org/10.1111/add.12405

Ort, C., Lawrence, M.G., Reungoat, J., Mueller, J.F., 2010a. Sampling for PPCPs in Wastewater Systems: Comparison of Different Sampling Modes and Optimization Strategies. Environ. Sci. Technol. 44, 6289–6296. http://dx.doi.org/10.1021/es100778d

Ort, C., Lawrence, M.G., Rieckermann, J., Joss, A., 2010b. Sampling for Pharmaceuticals and Personal Care Products (PPCPs) and Illicit Drugs in Wastewater Systems: Are Your Conclusions Valid? A Critical Review. Environ. Sci. Technol. 44, 6024–6035. http://dx.doi.org/10.1021/es100779n

Ort, C., van Nuijs, A.L.N., Berset, J.D., Bijlsma, L., Castiglioni, S., Covaci, A., de Voogt, P., Emke, E., Fatta-Kassinos, D., Griffiths, P., Hernández, F., González-Mariño, I., Grabic, R., Kasprzyk-Hordern, B., Mastroianni, N., Meierjohann, A., Nefau, T., Östman, M., Pico, Y., Racamonde, I., Reid, M., Slobodnik, J., Terzic, S., Thomaidis, N., Thomas, K. V., 2014b. Spatial differences and temporal changes in illicit drug use in Europe quantified by wastewater analysis. Addiction 109, 1338–1352. http://dx.doi.org/10.1111/add.12570

Simon, S.L., Richardson, K., Dacey, J., Glynn, S., Domier, C.P., Rawson, R.A., Ling, W., 2002. A comparison of patterns of methamphetamine and cocaine use. J. Addict. Dis. 21, 35–44.

van Nuijs, A.L.N., Pecceu, B., Theunis, L., Dubois, N., Charlier, C., Jorens, P.G., Bervoets, L., Blust, R., Neels, H., Covaci, A., 2009. Cocaine and metabolites in waste and surface water across Belgium. Environ. Pollut. 157, 123–129.

Zuccato, E., Chiabrando, C., Castiglioni, S., Calamari, D., Bagnati, R., Schiarea, S., Fanelli, R., 2005. Cocaine in surface waters: A new evidence-based tool to monitor community drug abuse. Environ. Heal. 4, 14–17. http://dx.doi.org/10.1186/1476-069X-4-14

S 9. Supplement

S 9.1. Limits of detection and quantification

The chemical analysis procedure established the detection and quantification limits to create a minimum signal-to-noise ratio of 3 and 10, respectively. This resulted in the limits presented in Table S.1. The amount of censoring for each municipality-drug combination is in Table S.2.

	Meth	MDMA	Benzoylecgonine	Methadone	Hydrocodone	Oxycodone
LOD	1.5	1.0	1.0	2.0	2.0	2.0
LOQ	10.0	2.5	10.0	2.5	2.5	2.5

Note: Meth = methamphetamine

Table S.1: Detection and quantification limits (ng/L)

S 9.2. Sampled WWTPs

The 19 treatment plants contributing data for this analysis represent a convenience sample of diverse systems in the Northwest United States, specifically in the states of Washington and Oregon. The given WWTP names are a mix of neighborhood names, city names, or collective names or otherwise convenient labels that may or may not accurately reflect the catchment area served. We use the term "municipality" in general to indicate the WWTP and the area served. While most municipalities are along the main north-south transportation corridor (US Interstate 5), several locations represent smaller cities or towns along the coast or inland in the drier regions east of the Cascade mountain range. The relative population sizes covered by each WWTP and their representative level of precipitation for 2009 are presented in Figure S.1. The final (average for Seattle and Renton; see Section 3.1) population estimates and the source of those estimates, per the WWTP surveys, are in the second and third columns of Table S.2.

S 9.3. Sample data setup

Calculating a given sample's index load requires four parameters: the measured concentration (column 1 in Table S.3), the titration correction factor to account for titrating the sample to facilitate measurement (column 2), the day's measured flow through the WWTP (column 3), and the day's population estimate (column 6, constant in this example; see Section 3.1 for discussion of how this value might vary). The "Final Concentration" and "Final Load" columns (4 and 5) in the sample data setup are intermediate calculations that represent the titration-corrected concentration and the product of this concentration and the day's volume.

For the third and fourth rows, the initial concentrations are censored. For estimating the mean index load in the presence of censoring via a survival analysis method, the censoring must be represented as intervals to indicate the range the values could take. The final two columns, here labeled "Left" and "Right" (your software may prefer "Begin" and "End", etc.), are what are inputted to the estimating software. For measurable observations, the two sides of this interval are the calculated index load. For values below the level of detection (<LOD, where the LOD in the example is 1.5), the left side of the interval is 0 and the right side is the LOD×titration factor×flow÷population÷(1000^2), to reflect that the initial concentration must be in the range of 0 to 1.5. For values below the level of quantification (<LOQ, where LOQ here is 10.0), we only know the initial concentration is between the LOD and the LOQ, and so the two sides of the interval reflect these values multiplied through: E.g., LOQ×titration factor×flow÷population÷(1000^2).

S 9.4. Robust retransformation

After maximum likelihood estimation, implemented with 50–80% censoring, we apply a robust retransformation to avoid retransformation bias as described by Helsel (2012), Kroll and Stedinger (1996), and Shumway, Azari, and Kayhanian (2002). In order to create the distribution graphs in Figures 1, S.2, and S.3, we also apply Steps 1–4 below for cases with less than 50% censoring. Given data as in the example above (Table S.3), the primary goal is to have an unbiased estimate of the mean of the data, and the

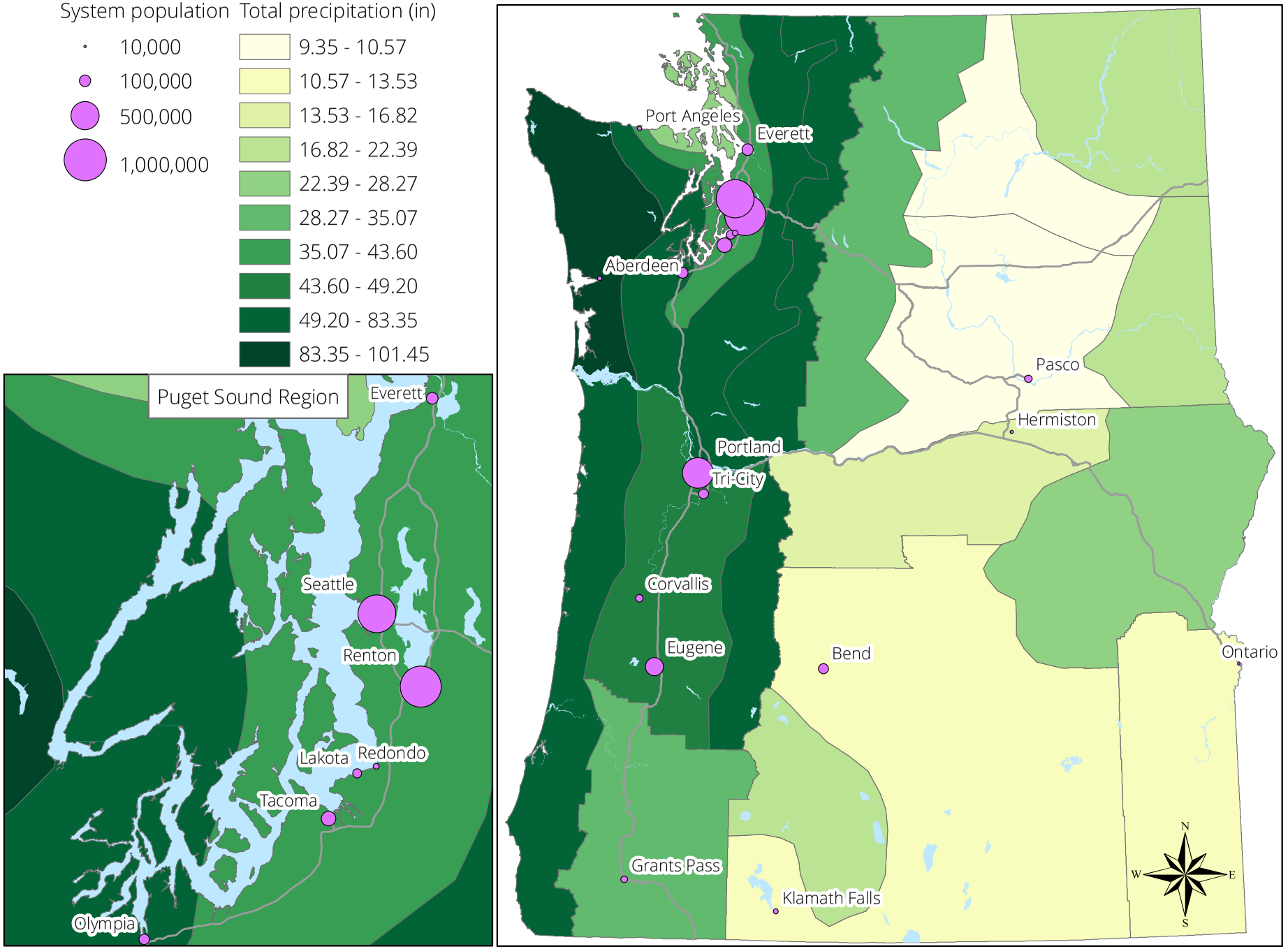

Figure S.1: WWTPs with estimated population and 2009 precipitation by National Oceanographic and Atmospheric Administration climate division

	Population estimate		Methamphetamine		MDMA		Benzoylecgonine		Methadone		Hydrocodone		Oxycodone	
WWTP	Source	Final	N	Censored	N	Censored	N	Censored	N	Censored	N	Censored	N	Censored
Aberdeen	Census	21170	42	0	43	81.4%	43	0	43	0	42	9.5%	42	2.4%
Bend	Census	85480	48	0	54	3.7%	54	0	54	0	54	1.9%	53	0
Corvallis	state	59562.56	55	0	52	55.8%	52	28.8%	53	22.6%	52	11.5%	53	7.5%
Eugene	NA#	234300	51	0	55	7.3%	55	3.6%	55	0	55	1.8%	55	0
Everett	NA#	126535.5	46	0	53	24.5%	52	0	52	0	52	0	52	0
Grants Pass	connections	46000	50	0	53	50.9%	44	0	53	0	54	0	54	0
Hermiston	state	16653.6	53	0	52	94.2%	52	0	52	0	52	0	52	0
Klamath Falls	Census	32852.87	52	0	52	51.9%	53	50.9%	52	0	52	0	53	0
Lakota	NA	73528	54	0	53	67.9%	53	0	53	0	53	1.9%	53	0
Olympia	connections	98000	55	0	52	57.7%	52	1.9%	52	0	52	30.8%	53	15.1%
Ontario	NA#	13516.9	43	0	55	92.7%	55	23.6%	55	0	55	0.0%	55	0
Pasco	state	53304.05	52	0	52	44.2%	52	0	52	7.7%	53	3.8%	51	2.0%
Port Angeles	Census	20844.1	53	0	55	72.7%	53	7.5%	53	1.9%	53	9.4%	53	1.9%
Portland	state	637842.5	54	0	43	32.6%	42	9.5%	40	0	43	11.6%	43	18.6%
Redondo	NA	35552	53	1.9%	48	43.8%	48	0	48	0	34	2.9%	48	0
Renton	NA	938535.9	50	0	46	47.8%	46	0	46	2.2%	46	19.6%	46	10.9%
Seattle	NA	841326.3	53	0	50	16.0%	50	0	50	0	50	20.0%	50	16.0%
Tacoma	NA#	178086.6	53	0	50	8.0%	50	0	50	0	50	4.0%	50	0
Tri-City	connections	87152	54	0	54	61.1%	54	0	54	0	54	0	54	0

Note: N is number of samples available for analysis. NA in population estimate source indicates missing source; NA# indicates Census-based daytime population adjustment used. Seattle and Renton populations indicate average across all sampled days; others populations constant.

Table S.2: Population estimates and censoring of samples by WWTP

1	2	3	4	5	6	7	8	9
Measured Concentration (ng/L)	Titration Correction Factor	Flow (L/day)	Final Concentration (ng/L)	Final Load (grams)	Population	Index Load (mg/person/day)	Left	Right
1005.2	1.0767	17261424	1082.3	18.682	21165	0.8827	0.8827	0.8827
1782.0	1.0538	11318346	1877.9	21.254	21165	1.0042	1.0042	1.0042
<LOD	1.0560	15375268			21165		0.0000	0.0012
<LOQ	1.0490	15568495			21165		0.0012	0.0077

Table S.3: Sample data setup for substance with LOD of 1.5 ng/L and LOQ of 10 ng/L

secondary goal to have a full hypothetical dataset to create boxplots. The robust retransformation is implemented as follows:

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

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

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

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.

S 9.5. Distributions and estimates for other analytes

Following the discussion of censoring, distribution, error components, and estimates for methamphetamine and MDMA in the main text, here we present the graphical results for the other analytes considered: benzoylecgonine, methadone, hydrocodone, and oxycodone. In the boxplot graphs, the shading of the box represents the amount of censoring for each municipality-drug combination, with darker boxes indicating more complete data. The WWTPs are ordered by the completeness of the data and then alphabetically. The background shading and label (Report censoring, MLE, KM, Complete data) indicate for groups of municipality-drugs what method was used to create the estimate in the corresponding estimates graphs. In the latter, the shading of the estimate indicator (the diamond) represents the amount of censoring, and the dotted line the unweighted average estimate across the 19 WWTPs. WWTPs are ordered by the mean annual estimate.

S 9.6. Sensitivity of confidence intervals to changes in U_S

As described in the Discussion, our annual sampling error estimate of $U_S = 10\%$ came from a study of sampling (or monitoring) uncertainty in a small European city (Ort et al., 2014a). That exercise involved repeated samples of $N = 56$ taken from over 1000 consecutive daily samples of benzoylecgonine (cocaine metabolite). With fewer than 56 samples in the current analysis, WWTP catchment area sizes of more or less than 7000 users, and substances that may have more or less variation in use than cocaine, one may question whether a different estimate of annual sampling error might change the results substantially. In Figure S.6, we present the estimates and confidence intervals for a single substance, MDMA, in which we have both doubled our sampling uncertainty RSD (top panel) and halved it (bottom panel). (Similar graphs for the other analytes are available upon request.) Compared to the bottom panel of Figure 2, we see that changing one of our four uncertainty parameters has small effects on the resulting CIs. With a 5% uncertainty, Port Angeles MDMA becomes more clearly significantly higher than average, but we remain uncertain that Renton MDMA could not be essentially 0. With less certainty about the ability of our 43 to 55 samples to represent the whole year—$U_S = 20\%$—Port Angeles MDMA levels are clearly not significantly different than the average, and Tacoma's CI is more likely to overlap with any other WWTP CI, but all except the Renton index load remain significantly greater than 0. Compared with both the size of the CIs and the differences in estimated means, these changes in CI due to different annual sampling uncertainties are relatively small.

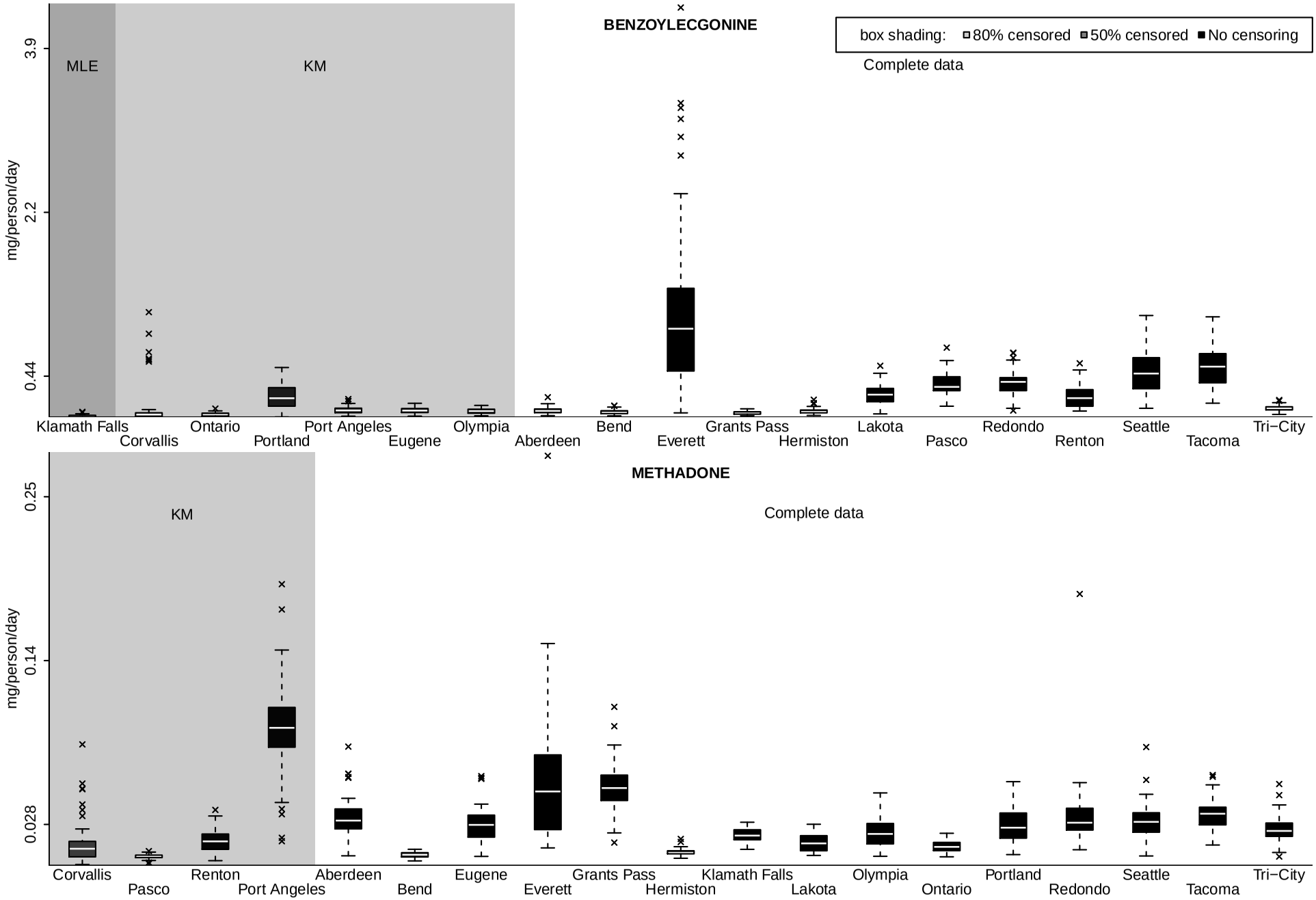

Figure S.2: Benzoylecgonine (top) and methadone (bottom) index load distribution (mg/person/day)

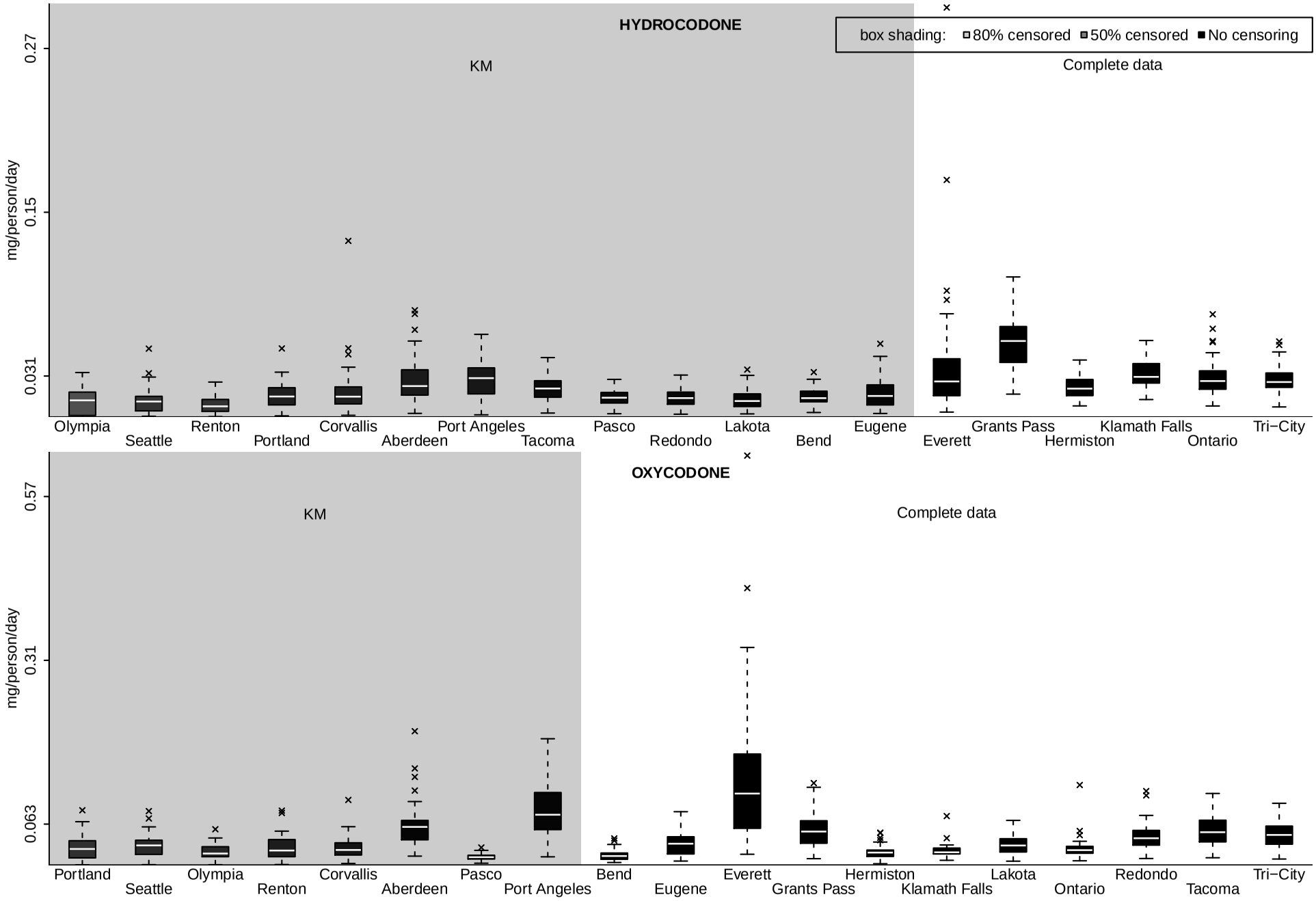

Figure S.3: Hydrocodone (top) and oxycodone (bottom) index load distribution (mg/person/day)

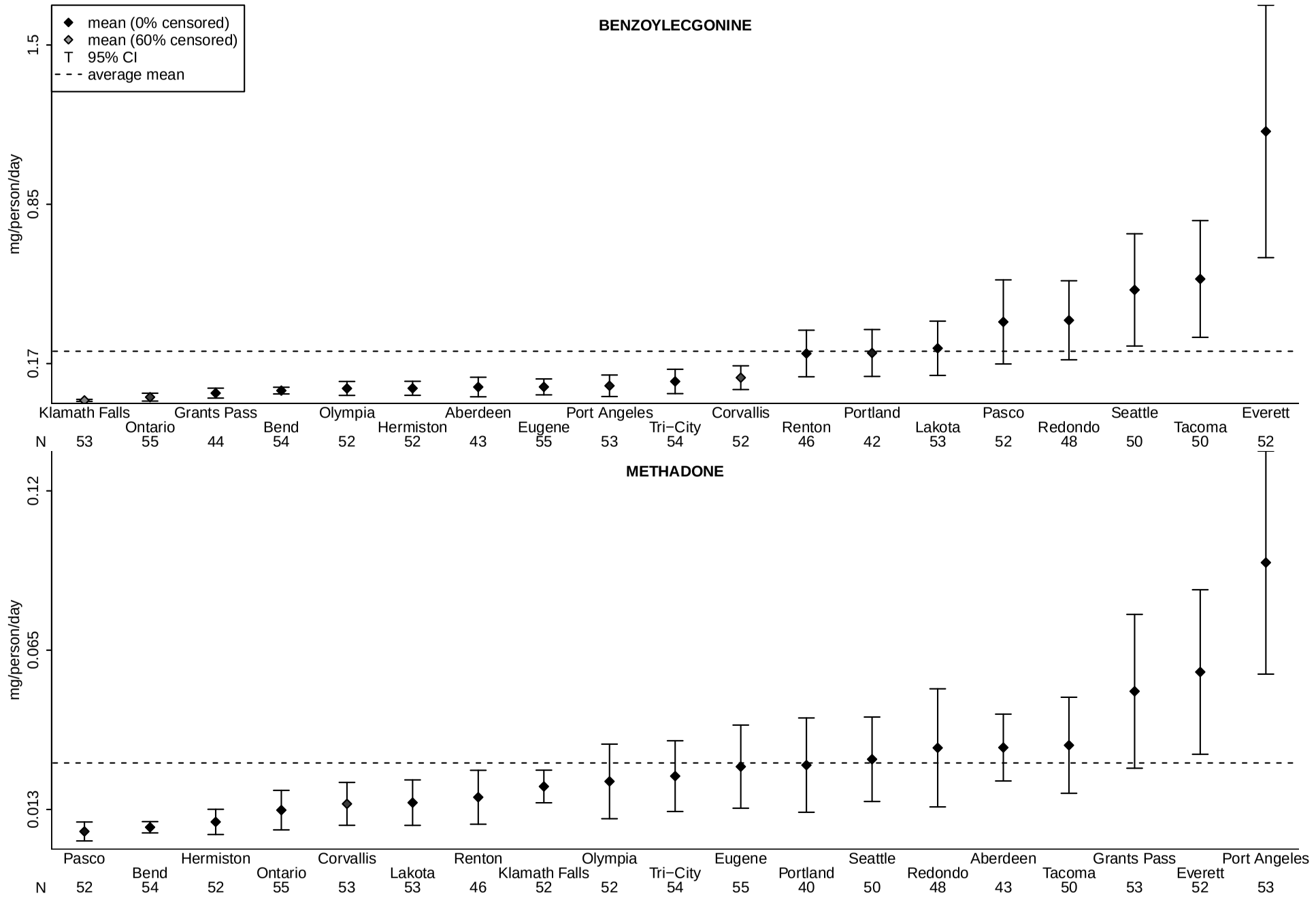

Figure S.4: Benzoylecgonine (top) and methadone (bottom) index load mean with 95% confidence interval (mg/person/day)

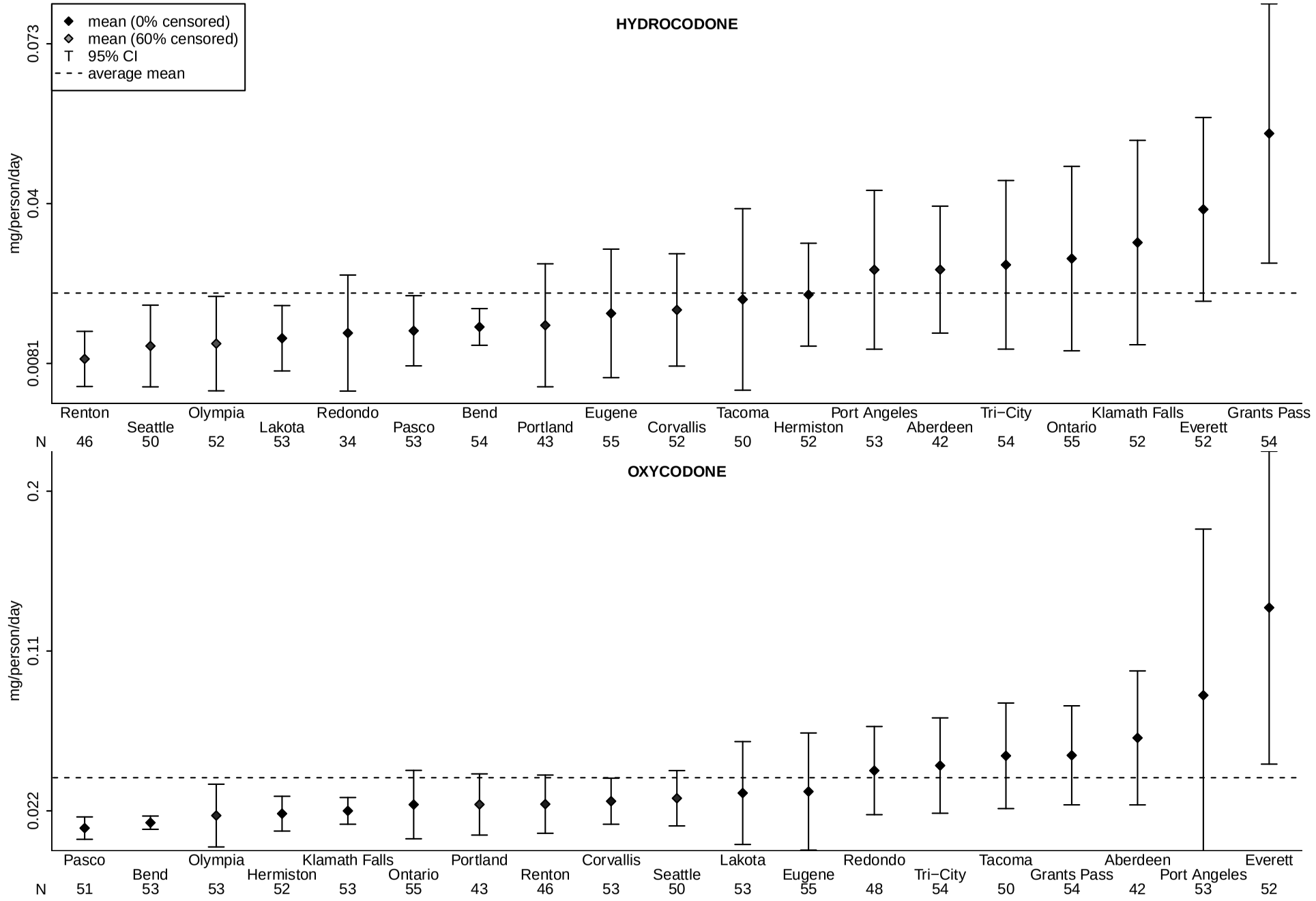

Figure S.5: Hydrocodone (top) and oxycodone (bottom) index load mean with 95% confidence interval (mg/person/day)

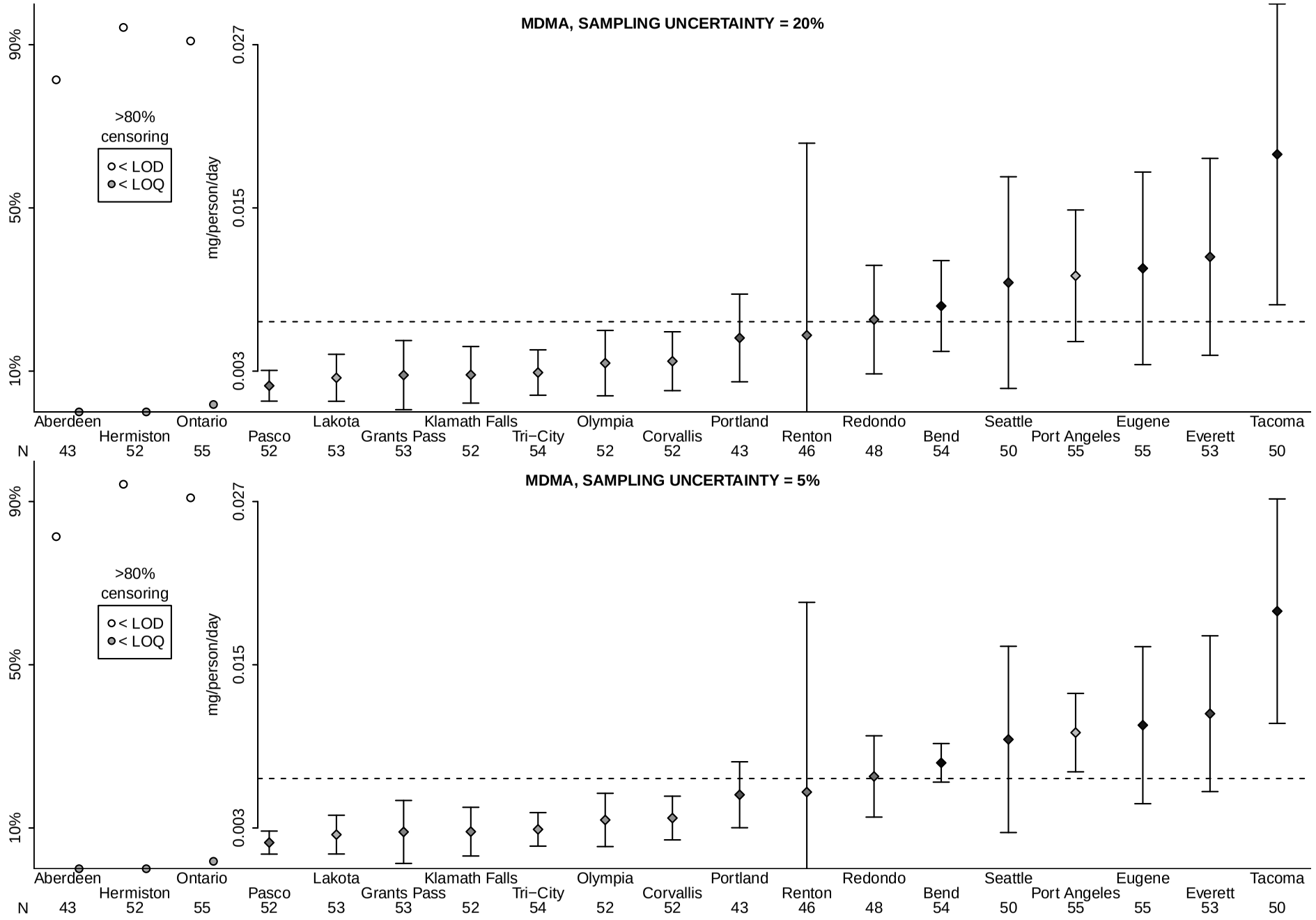

Figure S.6: MDMA index load mean with 95% confidence interval (mg/person/day) based on annual sampling error of 20% (top) and 5% (bottom)

S 9.7. Observation variability and population size

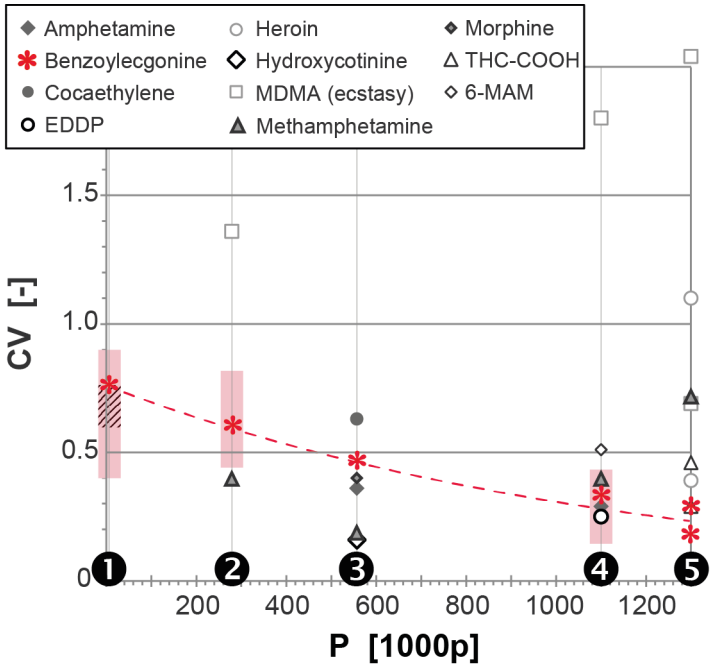

Figure S.7: Variability of daily drug loads [as coefficient of variation (CV)] vs. population size (P, in 1000s) in five different catchments for different substances (duration of studies in days: ①=1369, ②=311, ③=28, ④=239, ⑤=28/35). Reasons for CVs exceeding 0.8 in theses studies are: i) observations below limit of quantification (heroin) and ii) pronounced intra-week variability (MDMA). Such a regular weekend effect causes a high CV but does not imply that more samples would be needed for the same acceptable uncertainty. Adapted from European Monitoring Centre for Drugs and Drug Addiction (2016).