Science of the Total Environment xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

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

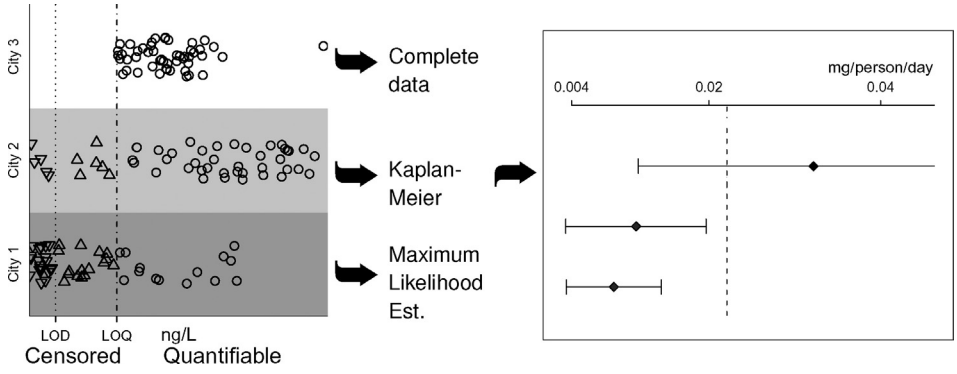
Caleb J. Banta-Green [a,*], Alex J. Brewer [b,c], Christoph Ort [d], Dennis R. Helsel [e], Jason R. Williams [a], Jennifer A. Field [b]

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

HIGHLIGHTS

- Wastewater entering treatment plants can be a population drug use data source.
- Careful consideration of the components of such an estimate is required.
- Censored observations are common and can be used to create an estimate.
- We demonstrate how to assemble the components to create estimates of drug levels.
- Results can appropriately be used to compare estimates across locations and drugs.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 22 March 2016
Received in revised form 8 June 2016
Accepted 9 June 2016
Available online xxxx

Keywords:
Wastewater
Drug epidemiology
Wastewater-based epidemiology
Censored data
Summary statistics
Data visualization

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.

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.

* Corresponding author.
E-mail address: calebbg@uw.edu (C.J. Banta-Green).

http://dx.doi.org/10.1016/j.scitotenv.2016.06.052
0048-9697/

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.

1. Introduction

Analyzing wastewater for drugs of abuse and medications dates from the 1970s (Hignite and 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 and 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

Please cite this article as: Banta-Green, C.J., et al., Using wastewater-based epidemiology to estimate drug consumption—Statistical analyses and data presentation, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.06.052