Science of the Total Environment 487 (2014) 642–650

Contents lists available at ScienceDirect

Science of the Total Environment

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

Illicit and pharmaceutical drug consumption estimated via wastewater analysis. Part B: Placing back-calculations in a formal statistical framework

Hayley E. Jones [a,*], Matthew Hickman [a], Barbara Kasprzyk-Hordern [b], Nicky J. Welton [a], David R. Baker [c], A.E. Ades [a]

[a] School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK
[b] Department of Chemistry, University of Bath, Faculty of Science, Bath BA2 7AY, UK
[c] Mass Spectrometry Business Unit, Shimadzu, Wharfside, Manchester M17 1GP, UK

HIGHLIGHTS

- Analysis of wastewater allows estimation of illicit drug consumption.
- However, it is crucial to formally acknowledge the many sources of uncertainty.
- The simple and flexible Monte Carlo simulation approach allows this.
- There are many software options: we provide an Excel spreadsheet and R code.
- Bayesian modelling using Markov chain Monte Carlo allows interesting extensions.

ARTICLE INFO

Article history:
Received 18 November 2013
Received in revised form 20 February 2014
Accepted 21 February 2014
Available online 15 March 2014

Keywords:
Sewage epidemiology
Monte Carlo simulation
Uncertainty propagation
Bayesian modelling
Illicit drugs

ABSTRACT

Concentrations of metabolites of illicit drugs in sewage water can be measured with great accuracy and precision, thanks to the development of sensitive and robust analytical methods. Based on assumptions about factors including the excretion profile of the parent drug, routes of administration and the number of individuals using the wastewater system, the level of consumption of a drug can be estimated from such measured concentrations. When presenting results from these 'back-calculations', the multiple sources of uncertainty are often discussed, but are not usually explicitly taken into account in the estimation process. In this paper we demonstrate how these calculations can be placed in a more formal statistical framework by assuming a distribution for each parameter involved, based on a review of the evidence underpinning it. Using a Monte Carlo simulations approach, it is then straightforward to propagate uncertainty in each parameter through the back-calculations, producing a distribution for instead of a single estimate of daily or average consumption. This can be summarised for example by a median and credible interval. To demonstrate this approach, we estimate cocaine consumption in a large urban UK population, using measured concentrations of two of its metabolites, benzoylecgonine and norbenzoylecgonine. We also demonstrate a more sophisticated analysis, implemented within a Bayesian statistical framework using Markov chain Monte Carlo simulation. Our model allows the two metabolites to simultaneously inform estimates of daily cocaine consumption and explicitly allows for variability between days. After accounting for this variability, the resulting credible interval for average daily consumption is appropriately wider, representing additional uncertainty. We discuss possibilities for extensions to the model, and whether analysis of wastewater samples has potential to contribute to a prevalence model for illicit drug use.

1. Introduction

The analysis of communal sewage water entering wastewater treatment plants (WWTPs) offers potential for enhancing our knowledge of illicit drug consumption (Daughton, 2001; Frost et al., 2008; van Nuijs et al., 2011a; Zuccato et al., 2008). State-of-the-art sensitive and robust analytical methods mean that concentrations of drug target residues (DTRs), such as metabolites of an illicit drug, in wastewater can be measured with great accuracy and precision (Baker and Kasprzyk-Hordern, 2011b; Castiglioni et al., 2013). In what has been termed the 'sewage epidemiology' approach, consumption of the parent drug is 'back-calculated'

* Corresponding author. Tel.: +44 117 331 3941; fax: +44 117 928 7325.
E-mail address: hayley.jones@bristol.ac.uk (H.E. Jones).

http://dx.doi.org/10.1016/j.scitotenv.2014.02.101

from these measured DTR concentrations (Zuccato et al., 2005). For example, our sister paper, entitled 'Illicit and pharmaceutical drug consumption estimated via wastewater analysis. Part A: Chemical analysis and drug use estimates' (Baker et al.,2014–in this issue), describes a study of a large UK population.

The results of these calculations are of course only estimates of illicit drug use, subject to many sources of uncertainty. As the field develops, it is important that this is properly addressed. Key variables include the size of the population served by the WWTP and the percentage of a dose of the parent drug that is excreted as the DTR. In addition, there is evidence that this percentage varies according to the route of administration of the parent drug (Khan and Nicell, 2011). Hence data are also required on the distribution of routes of administration across the population. All parameters informed by data are subject to sampling variation.

Usually, as in Baker et al. (2014–in this issue), only the analytical uncertainty in the measurement of DTR concentrations in a wastewater sample has been explicitly taken into account. Since this uncertainty is generally very small, back-calculated drug consumption estimates often incorrectly appear to be very precise. To avoid over-interpretation of the estimates, it is highly desirable to present credible intervals around them, accounting for as many additional sources of uncertainty as possible. Recently, Lai et al. (2011) and Mathieu et al. (2011) have attempted to propagate uncertainty in multiple parameters simultaneously through the back-calculations. However, as we will discuss below, their approach has limited applicability.

In this paper we propose Monte Carlo simulation as a simple and more general approach to accounting for multiple sources of uncertainty in sewage epidemiology back-calculations. The approach involves specifying a probability distribution, based on 'all available evidence', for each of the parameters involved. The specified distributions are repeatedly sampled from at random, and the back-calculations performed for each set of simulated values. The end result is a simulated distribution for consumption of the parent drug, from which summary statistics can be presented which appropriately reflect the uncertainties. Monte Carlo simulation has been routinely used to propagate uncertainty in models in the physical and social sciences since use of computers became widespread (Metropolis and Ulam, 1949). It also has a key role in decision making (Critchfield and Willard, 1986; Doubilet et al., 1985), as it provides a simple way of estimating expectations under uncertainty in non-linear models. To demonstrate its application to wastewater analysis, we use data from the Part A paper (Baker et al.,2014–in this issue) to 'back-calculate' cocaine consumption based on concentrations of the metabolites benzoylecgonine and norbenzoylecgonine (Section 3).

A further possibility, which we illustrate in Section 4, is simulation from a Bayesian joint posterior distribution using Markov chain Monte Carlo (Gilks et al., 1996). This has the advantage of combining simulation with statistical estimation of parameters from multiple data sources. This approach – sometimes called 'comprehensive decision analysis' – has been popular in decision sciences for over 30 years (Parmigiani, 2002; Samsa et al., 1999; Spiegelhalter et al., 1999). For wastewater analysis, it opens up possibilities for many more sophisticated statistical analyses, such as modelling variability over time or allowing consumption of a drug to be simultaneously informed by concentrations of multiple DTRs.

2. Background: 'back-calculation' of drug consumption using DTR concentrations

Baker et al. (2014–in this issue) present estimates of drug and pharmaceutical consumption in a large (estimated 3.4 million) urban UK population. They used the following modified versions of formulae introduced by Zuccato et al. (2005) to estimate per capita consumption from measured DTR concentrations:

Load of DTR in grams grammes per day

$$Load = \frac{Concentration \times Flow}{1000} \times \left(\frac{100}{100 + Stability}\right) \times \left(\frac{100}{100 - Sorption}\right) \quad (1)$$

where*Concentration* = DTR concentration in wastewater influent (ng/l),*Flow* = volume of flow to the wastewater influent over a 24 hour period (millions of litres/day),*Stability* = percentage change in concentration of the DTR in wastewater in the conditions (time, pH and temperature) relevant to the study, and *Sorption* = percentage sorption of the DTR to suspended particulate matter (SPM) in wastewater.

Estimated drug consumption in mg/day per 1000 people

$$Consumption = \left(\frac{Load}{Population \times Excretion}\right)\left(\frac{MW_{Par}}{MW_{DTR}}\right) - OS \quad (2)$$

where *Excretion* = proportion of a dose of the parent drug excreted as the DTR, MW_{Par} = molecular weight of the parent compound, MW_{DTR} = molecular weight of the DTR, *Population* = size of the population served by the WWTP (millions), and *OS* = the amount of the DTR present in wastewater due to sources other than consumption of the parent compound (e.g. hospital or prescription usage).

For drugs such as cocaine that are administered using multiple routes by different users, the typical metabolism profile of the drug will likely vary according to this. As such, *Excretion* should be estimated as an average over the different routes (Khan and Nicell, 2011):

Proportion of a dose of the parent drug excreted as the DTR

$$Excretion = \Sigma_R \Big[(\text{proportion of all parent drug mass that is administered by route } R) \times (\text{proportion of a dose of the parent drug excreted as the DTR following administration by route } R)\Big] \quad (3)$$

Except for the molecular weights, there is uncertainty about all of these parameters. Failure to take these uncertainties into account is likely to lead to over-interpretation of the results.

When uncertainty about the individual parameter values involved in the back-calculations has been quantified, it has generally been expressed as relative standard deviations (SD) (Castiglioni et al., 2013). The RSD is defined as the standard deviation divided by the absolute value of the parameter estimate. We note that there is ambiguity here in the meaning of 'standard deviation'. Consider, for example, the *Excretion* factors in Eq. (3). Clearly the metabolism profile of a drug will vary across individuals, according for example to genetic factors. This variability is quantified by the standard deviation. But for valid inference on consumption by a large population, only the *average* excretion profile across the population of users need be well estimated. The standard deviation of a parameter estimate is usually called the 'standard error' (SE) in statistics. The SE is the more appropriate measurement of uncertainty about the parameter used in the back-calculation. It can be reduced by the collection of new data, whereas the standard deviation (SD) cannot. In the simple case where a parameter has been estimated by the arithmetic mean of n data points, the SE is calculated as $SD/\sqrt{n}$. When the parameter estimate is a weighted average of estimates across multiple studies, then it is the standard error of the pooled estimate that we generally require.

If the formulae for estimating consumption were linear on the log scale, then the square of the RSD of the estimate of consumption could be approximated by the sum of squares of the RSDs of each individual parameter estimate (Lai et al., 2011; see also Mathieu et al. 2011 who used this approach to quantify uncertainty in estimated loads). This may have been a reasonable approach for early back-calculations, when *Stability*, *Sorption* and *OS* were not accounted for and *Excretion* was estimated by a single value rather than by averaging across routes