Effects of tap water processing on the concentration of disinfection by-products

APPENDIX

Statistical treatment of data below a detection limit when the sample size is very small

Values that are below the detection limit (BDL) for an assay of an environmental agent in air, water or food are common in exposure measurement in environmental epidemiology (Helsel 2005). They present a problem when calculating summary statistics for, or testing hypotheses using, the exposure measurements.

To date, at least 14 different methods have been proposed to deal with this issue (Zhang et al. 2009). They can be broadly categorised into 6 groups: (1) simple deletion of the BDL values (Zhang et al. 2009); (2) substitution methods, entailing substitution of a single value for all the BDL values or substitution from a distribution (Lin & Niu 1998); (3) distributional methods, where the characteristics of an assumed distribution of BDL values are used to calculate the summary statistics, data below and above the detection limit are assumed to follow the distribution and maximum likelihood methods are used to estimate the distributional parameters (Helsel & Hirsch 2002); (4) robust methods, in which values for the BDL data are extrapolated from a distribution fitted to the observed data (Helsel & Hirsch 2002); (5) multiple imputation methods, where the BDL values are imputed from the association between the observed values and some other covariates (Lubin et al. 2004); and (6) nonparametric methods that do not require the assumption of any distribution and give the same rank to all the BDL values for hypothesis testing (Helsel 2005; Clarke 1998).

Only the substitution method is applicable when an estimate is required for each BDL value (as, for example, when making paired comparisons), when the sample size is too small to assume any distribution (e.g. <5) and all or all but one of the observations in one or both of the paired sets of observations have BDL values. Small sample properties for two independent samples have been well established (Clarke 1998; Gilliom & Helsel 1986), but there has been less exploration of the small sample properties for paired data with censoring.

There are, however, problems with substitution methods. When single substitution is used, the p-value from a hypothesis test varies with the value substituted in an unpredictable way (Helsel 2005). In addition, single substitution reduces the sample variance and thus biases estimates of precision in the parameter estimates. The method of substitution from a distribution provides a partial solution to this problem (Lin & Niu 1998). For this approach, it is generally assumed that the BDL values are uniformly distributed between zero and the detection limit. A set of random values are generated from this distribution to substitute for the BDL values.

We conducted a field survey in Sydney to estimate the concentration of different species of disinfection by-products (DBPs) in drinking water and to evaluate the effect on DBP concentrations of different ways of treating drinking water (e.g., boiling, filtering or storing in a refrigerator). By log transforming the data, we calculated geometric mean concentrations of different DBP species with their 95% confidence intervals and compared the geometric mean for each species after each treatment method with that with no treatment and estimated p-values using a paired t-test. Sample sizes were very small (4 or 5) for some species by treatment categories and BDL values (usually 1 to 3) were common in each category.

To analyse the data we first used the simple substitution method and substituted two-thirds of the detection limit for each BDL value (Whitaker 2005). Confidence intervals were unrealistically narrow or could not be calculated (when all values were missing). We then tried the method of substitution from a distribution by sampling the BDL values from a uniform distribution ranging from 0 to 1 (where 1 is the detection limit) with an arithmetic mean of 0.5 and a geometric mean of 0.37. However, since the number of BDL values was small, the mean of the substituted observations was rarely equal to the mean of the distribution and replication of the method gave very different estimates.
To deal with this issue we repeated the whole process (calculation of means and their confidence intervals, and the paired $t$-test) 10,000 times by simulation. The simulated estimates were normally distributed and the means of the simulated means, confidence intervals and the $p$-values were taken as the final estimates. Simulation output for one of the species is given in Table A1.

The lower detection limit (LDL) may vary by measure, For individual $i$ with actual value $y_i$, let the observed value $y_i^{\text{observed}}$ be $y_i$ if $y_i > \text{LDL}$ otherwise let $y_i^{\text{observed}} = 0$. For each simulation, we substituted the observed values less than the LDL with a random realisation that was uniform between 0 and the LDL, such that $y_i^{\text{simulated}} \sim \text{Uniform}(0, \text{LDL})$ if $y_i^{\text{observed}} < \text{LDL}$, otherwise $y_i^{\text{simulated}} = y_i^{\text{observed}}$. We then calculated means, confidence intervals and $t$-tests on log ($y_i^{\text{simulated}}$) for all of the individuals, performed the simulation 10,000 times, and averaged the summary statistics across the simulations.

**REFERENCES**


Whitaker, H. 2005 Exposure Assessment of Chlorination Disinfection by-products for use in Epidemiological Studies of Water Quality and Birth Outcomes. PhD, Imperial College, School of Medicine.

---

**Table A1** | Output of simulation for TCA in instant boiling water where 9 of 11 observations had BDL values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean (95% confidence bound)$^a$</th>
<th>Difference from mean TCA concentration in untreated water (95% confidence bound)$^b$</th>
<th>$p$ value$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of simulated values</td>
<td>0.44 (0.24, 0.82)</td>
<td>15.18 (8.09, 28.09)</td>
<td>0.0000187</td>
</tr>
<tr>
<td>Minimum simulated value</td>
<td>0.16 (0.04, 0.43)</td>
<td>7.92 (4.95, 9.02)</td>
<td>7.56e–12</td>
</tr>
<tr>
<td>Maximum simulated value</td>
<td>0.85 (0.77, 1.33)</td>
<td>40.83 (15.52, 177.65)</td>
<td>0.0014469</td>
</tr>
</tbody>
</table>

$^a$Geometric mean; all calculations were done in log scale and then back transformed for reporting.

$^b$There was no BDL value for the TCA concentration in untreated water.

$^c$p-value from the paired $t$-test.