Estimation of Individual Rodent Water Consumption from Group Consumption Data for Gestation, Lactation, and Postweaning Life Stages Using Linear Regression Models

Elizabeth J. Slone, Linda K. Teuschler, Laura L. Aume, Michael G. Narotsky, Jane Ellen Simmons, Robert A. Lordo, and Glenn E. Rice

Elizabeth J. Slone, M.S., is a statistician who conducted this work while at Battelle, Columbus, OH; Linda K. Teuschler, M.S., is a statistician at NCEA/ORD/U.S. EPA, Cincinnati, OH; Laura L. Aume, M.S., is a statistician at Battelle, Columbus; Michael G. Narotsky, Ph.D., is a toxicologist at NHEERL/ORD/U.S. EPA, Research Triangle Park, NC; Jane Ellen Simmons, Ph.D., is a toxicologist, also at NHEERL/ORD/U.S. EPA, Research Triangle Park; Robert A. Lordo, Ph.D., is a statistician at Battelle, Columbus; and Glenn E. Rice, Sc.D., is an environmental health scientist at NCEA/ORD/U.S. EPA, Cincinnati, OH.

Address correspondence and reprint requests to Linda K. Teuschler, U.S. EPA/NCEA-Cin, 26 W. ML King Dr. (MSA-110), Cincinnati, OH 45268, telephone 513-569-7573, or email teuschler.linda@epa.gov.

Abbreviations that appear ≥3× throughout this article: ADWC, average daily water consumption; AIC, Akaike Information Criteria; DBP, disinfection by-product; EPA, Environmental Protection Agency.

Abstract

In rodent bioassays where chemicals are administered in the drinking water, water consumption data for individual animals are needed to estimate chemical exposures accurately. If multiple animals share a common water source, as occurs in some studies, only the total amount of drinking water consumed by all animals utilizing the common source is directly measurable, and water consumption rates for individual animals are not available. In the Four Lab Study of the US Environmental Protection Agency, which included a multigenerational rodent bioassay, a complex mixture of drinking water disinfection by-products was delivered to multiple Sprague-Dawley rats from a common drinking water container. To estimate disinfection by-product mixture exposure for each animal, authors developed four log-linear regression models to allocate water consumption among rats sharing a common water container. The four models represented three animal lifestages: Gestation, Lactation, and Postweaning, with separate Postweaning models for male and female. Authors used data from six Sprague-Dawley rat bioassays to develop these models from available individual cage data for the Postweaning models, and available individual animal data for the Gestation and Lactation models. The $r^2$ values for the model fits were good, ranging from 0.67 to 0.92. The Gestation and Lactation models were generally quite accurate in predicting average daily water consumption whereas the Postweaning models were less robust. These models can be generalized for use in other reproductive and developmental bioassays where common water sources are used and data on the explanatory variables are available.

Key words: drinking water disinfection by-products (DBPs); Four Lab Study; regression analysis; reproductive toxicity; rodent drinking water consumption model

Introduction

In rodent bioassays that administer chemicals in the drinking water, water consumption data for individual animals are necessary to estimate chemical exposure accurately. In its reproductive toxicity guidelines, the U.S. Environmental Protection Agency (EPA) recommends recording water consumption data for bioassays where exposure occurs via drinking water (EPA 1996). Typically, water consumption is determined in rodent bioassays by weighing water bottles at specified intervals. When multiple animals share a common water source, as occurs in many studies, only the total amount of drinking water consumed by all animals utilizing the water source is directly measurable. Individual animal water consumption rates cannot be determined directly and typically are estimated by dividing the total amount of water consumed by the number of animals sharing the common water source.
Although water bottles are commonly used, they have several limitations. Exposure to volatile chemicals via water consumption may vary with the amount of water (and headspace) in the bottle because volatile chemicals are released into the headspace within the water bottle. For heat-sensitive chemicals, properly controlling water bottle temperature can be challenging. Water bottle use can generate large quantities of wasted test water as bottles are refreshed. Chemicals can absorb/adsorb to the surfaces of the bottle unless the system is constructed of polytetrafluoroethylene (e.g., Teflon®), stainless steel or glass, with glass presenting a potential safety hazard.

In view of the water bottle limitations described above, an alternative watering system was designed (McDonald et al. 2010) for the multi-generational rodent bioassay component of the EPA Four Lab Study, which evaluated a complex mixture of drinking water disinfection by-products (Narotsky et al. 2008b). The novel watering system included custom-made Teflon® bags that served as a single water source to all animals housed in up to five cages in the Four Lab Study, with each cage housing either a single gestating dam, a single lactating dam with her litter, or two animals postweaning. This system reduced interactions of the chemicals in the complex mixture with the components of the watering system, resulted in minimal headspace, decreased test water wastage, and provided a convenient way to maintain the water in a chilled environment that was protected from light (McDonald et al. 2010). In addition, the system reduced the burden associated with animal care. One drawback of the system is that individual animal water consumption rates were not measured.

To assess disinfection by-product (DBP) exposures in conjunction with analytical chemistry data measured during the multigenerational bioassay study known as the Four Lab Study (Dingus et al. 2011; Pressman et al. 2010; Simmons et al. 2002), it was necessary to estimate the amount of water consumed by individual study animals. Therefore, we developed the methods described in this article for estimating the water consumption of individual animals during a given time interval as a function of a number of predictor variables typically recorded in rodent reproductive bioassays and the total amount of water consumed among animals sharing a common water source.

This text presents a linear regression approach for estimating average daily water consumption rates (g/day) over a water consumption time interval for individual Sprague-Dawley rats. We developed four separate models to predict water consumption rates during three animal lifestages: Gestation, Lactation, and Postweaning, with separate Postweaning models for male and female. The models predict average daily water consumption (ADWC) rates for days within a water consumption time interval (ADWC), where the total amount of water consumed by the animals sharing the common water source is measured at the end of the interval. These predictions can be used to generate chemical exposure estimates that might be associated with specific days of gestation, lactation, or postweaning development.

Methods

Data Sources

To develop the linear regression models for predicting ADWC in individual study animals, we relied on information collected in six data sets (labeled A through F, Table 1). These data sets are from studies in the Four Lab Study that preceded the multigenerational bioassay. Data sets A, B, and D (Narotsky et al. 2008a, 2012) are from studies on water concentrates containing complex mixtures of DBPs. Data set C (Narotsky et al. 2012) is from a dose-range finding study with water containing sulfate and sodium, and Data sets E and F (Narotsky et al. 2010) are from a multigenerational study with a defined mixture of DBPs. In all of the studies, we used Sprague-Dawley rats aged 10 to 14 weeks at the start of gestation. The rats were housed in polycarbonate cages with perforated steel tops (Research Equipment Allentown Inc., Allentown NJ, Model IPC 10198URT30, and Lab Products Inc., Seaford DE, Model See-I) and heat-treated pine shavings as bedding. We administered drinking water via water bottle and collected individual water consumption data for each gestating female and lactating female (with her litter). For Postweaning animals (Data set F), we estimated average water consumption per animal for the one to three animals housed in each cage. For some litters, we maintained four males in two cages of two males each; however, for the current analyses, we regarded the samples from different cages as independent. In all of the bioassays, the controls received purified drinking water, and the treatment groups received different concentrations of DBPs (Table 1).

Table 1. Information on six water consumption data sets from Sprague-Dawley rat bioassays used in the model development process. a,b
| Data set A | Female rats, gestation  
| 3 dose groups (including 1 control): \( N = 20 \) per group  
| Treatment groups received either a concentrate prepared from chlorinated water or a concentrate prepared from ozonated/postchlorinated water  
| 13 water consumption intervals occurring during gestation and up to the date of parturition, each interval being 1 day in duration (from gestation days 8 through 20)  
| Narotsky et al. 2008a |
| Data set B | Female rats, gestation and lactation  
| 1 control group: \( N = 36 \); 1 treatment group: \( N = 35 \)  
| The treatment group received a concentrate prepared from chlorinated water  
| 2 water consumption intervals: one lasting 7 days (ending on gestation day 13), and the other lasting 3 days (ending on gestation day 16)  
| An additional 5-day water consumption interval occurred for animals achieving parturition on study day 22 or 23 (ending on gestation day 21)  
| Interpolation of body weight data was necessary when body-weights were not measured at the end of a given water consumption interval  
| Narotsky et al. 2012 |
| Data set C | Female rats, gestation and lactation  
| 1 control group: \( N = 5 \); 5 treatment groups: \( N = 11 \) or \( 12 \) per group  
| Treatment groups received water with varying levels of sodium and sulfate  
| 3 water consumption intervals occurring during gestation  
| 1 lasting 7 days (ending on gestation day 12)  
| 2 lasting 4 days (ending on gestation days 16 and 20)  
| 6 water consumption intervals occurring during lactation, lasting from 1 to 4 days (ending on postnatal days 8, 12, 16, 18, 20, and 21)  
| Three treatment groups had no body-weight data after postnatal day 8, and therefore, their data were not included in the model fitting for the Lactation model  
| Narotsky et al. 2012 |
| Data set D | Female rats, gestation and lactation  
| 1 control group: \( N = 11 \); 1 treatment group: \( N = 20 \)  
| Treatment group received chlorinated water concentrate  
| 4 water consumption intervals occurring during gestation, lasting from 2 to 4 days (ending on gestation days 11, 15, 18, 20)  
| 5 water consumption intervals occurring during lactation, lasting from 1 to 5 days (ending on postnatal days 4, 6, 8, 13, and 14)  
| 10 animals were removed after postnatal day 8  
| Narotsky et al. 2012 |
| Data set E | Female rats, gestation and lactation  
| 1 control and 3 treatment groups: \( N = 25 \) per group  
| Treatment groups received DBP mixtures composed of varying levels of four trihalomethanes and five haloacetic acids  
| 6 water consumption intervals occurring during gestation, ranging from 2 to 4 days (ending on gestation days 2, 6, 9, 13, 16, 20)  
| 7 water consumption intervals occurring during lactation, each lasting either 3 or 4 days (ending on postnatal days ranging from 5 to 26)  
| Data from the group treated with the highest water contaminant concentration were excluded because contaminant concentrations decreased water consumption relative to all other treatment groups |
Narotsky et al. 2010

Data set F
- Originated from same study as Data set E
- Male and female rats, postweaning; taken from litters represented in Data set E on postnatal day 26 and maintained through puberty
- 1 control and 3 treatment groups (N = 24 for females, 1 cage per litter; N = 34 to 35 for males, 1-2 cages per litter)
- Treatment groups received DBP mixtures composed of varying levels of four trihalomethanes and five haloacetic acids
- 12 water consumption intervals ranging from 3 to 4 days each (ending on postnatal days ranging from 29 to 69)
- 1 to 3 rats of a given sex were housed per cage. Each animal’s average daily water consumption was estimated by dividing the observed water consumption by the number of rats in the cage
- Data from the treatment group associated with the highest dose were excluded because water consumption was lower in this relative to in all other treatment groups
- Narotsky et al. 2010

*See text for references.

bDBP, disinfection by-product.

We used the following data sets to develop the different models: Data sets A through E to develop the Gestation model, B through E for the Lactation model, and F for the Postweaning models for both males and females. Table 2 provides a list of the candidate covariates (predictor variables) considered for predicting ADWC_i in the regression models. We identified candidate covariates based on knowledge of the important contributors to water consumption and the available data collected in the data sets. Because considerable collinearity, or correlation, existed between some of the candidate covariates, we could not clearly discern the effect of any one covariate on predicting ADWC_i within a model containing multiple covariates as predictor variables. Furthermore, the presence of some covariates in the model may imply that other covariates provide minimal additional predictive information and could therefore be excluded from the model. Thus, we designed the model development process to select the smallest subset of covariates whose predictive power would be statistically equivalent to a model containing a larger number of these covariates.

Table 2. Covariates considered for inclusion in final models to predict log-transformed average daily water consumption interval*a

<table>
<thead>
<tr>
<th></th>
<th>Gestation Model</th>
<th>Lactation Model</th>
<th>Female Postweaning Model</th>
<th>Male Postweaning Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Dam body weight (g) at end of water consumption interval</td>
<td>*Dam body weight (g) at end of water consumption interval</td>
<td>*Average rat body weight (g) at end of water consumption interval</td>
<td>*Average rat body weight (g) at end of water consumption interval</td>
<td></td>
</tr>
<tr>
<td>*Average daily measured change in dam body weight (g) during the interval</td>
<td>*Average daily measured change in dam body weight (g) during the interval</td>
<td>*Average daily measured change in average rat body weight (g) during the interval</td>
<td>*Average daily measured change in average rat body weight (g) during the interval</td>
<td></td>
</tr>
<tr>
<td>*Average daily percent change in dam body weight during the interval</td>
<td>*Average daily percent change in dam body weight during the interval</td>
<td>*Average daily percent change in average rat body weight during the interval</td>
<td>Average daily percent change in average rat body weight during the interval</td>
<td></td>
</tr>
<tr>
<td>*Water consumption interval length (# days)</td>
<td>*Water consumption interval length (# days)</td>
<td>Water consumption interval length (# days)</td>
<td>**Water consumption interval length (# days)</td>
<td></td>
</tr>
<tr>
<td>*Sodium concentration in water (g/L)</td>
<td>*Sodium concentration in water (g/L)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aSee text for references.

bDBP, disinfection by-product.
**Gestation Model**

<table>
<thead>
<tr>
<th>Sulfate concentration in water (g/L)</th>
<th>*Sulfate concentration in water (g/L)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>*Number of live and dead pups recorded during initial litter examination</th>
<th>*Number of live pups recorded during most recent litter examination</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>*Total litter weight recorded during initial litter examination</th>
<th>*Total litter weight recorded during most recent litter examination</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>*Number of rats in a cage</th>
<th>*Number of rats in a cage</th>
</tr>
</thead>
</table>

*Significant covariate in bivariate models, \( p \leq 0.10 \).

Sodium and sulfate concentrations in water were not considered in the Postweaning models because these concentrations were negligible in the water consumed by treatment groups within Data set F, the only study used for postweaning animals. Initial analysis of Data set C suggested that either sodium or sulfate concentration (for which data were available in the other five studies) could have a statistically significant effect on daily water consumption in gestating and lactating animals. Therefore, both concentrations were considered as candidate covariates for the Gestation and Lactation models.

**Model Development**

Initial data investigations suggested that water consumption rates are well approximated by a log linear (ln-linear) function of candidate covariates. Therefore, we applied the following ln-linear model to water consumption data from the studies in Table 1:

\[
Y_{ijkt} = \mu + \beta \cdot Time_{ijkt} + \sum_{m=1}^{m} \beta_{p(ij)k} \cdot X_{p(ij)k} + \epsilon_{ijkt},
\]

where

- \( Y_{ijkt} \) is the natural logarithmic-transformed average daily water consumption (ADWC) for the \( k^{th} \) rat in the \( f^{th} \) study in the \( i^{th} \) water treatment group for the water consumption interval ending on the \( t^{th} \) day,
- \( \mu \) is the intercept term, representing overall mean logarithmic transformed water consumption,
- \( \beta \) is the common slope parameter describing the average rate at which logarithmic-transformed water consumption measurements change per unit time,
- \( Time_{ijkt} \) is the time point corresponding to the \( t^{th} \) day when the water consumption measurement interval ended for the \( i^{th} \) water treatment group in the \( j^{th} \) study,
- \( \beta_{p(ij)k} (p=1, \ldots, m) \) are the slope parameters associated with the \( m \) selected covariates specified in Table 3 (for some integer \( m > 0 \)),
- \( X_{p(ij)k} \) are the values of the \( m \) selected covariates that are associated with the \( k^{th} \) rat in the \( i^{th} \) water treatment group in the \( j^{th} \) study,
\( \nu_i \) is the random effect associated with the \( i^{th} \) treatment group,

\( \alpha_{k(i)} \) is the random effect associated with the \( k^{th} \) rat within the \( i^{th} \) treatment group in the \( j^{th} \) study, and

\( \varepsilon_{t(ijk)} \) is residual error.

Model (1) is a mixed effect model in that it includes both fixed and random effect terms. The applied treatment was a random effect in the model because the definition of the treatments varied between studies, and the treatments were considered a sampling of all possible treatments that could have been administered. The term \( \varepsilon_{t(ijk)} \) represents residual error that is left unexplained by the model; for a given rat, the vector of residual errors representing water consumption levels at different time intervals \( t \) is assumed to have a spatial power covariance structure. Such a covariance structure allows for the possibility that a nonzero correlation may exist among the residual errors associated with the same rat among the time intervals and that errors for intervals that occur closer in time may have a higher correlation than errors for other intervals. The spatial power covariance structure differs from a first order autoregressive structure (AR(1)) in that the time intervals do not have to be equally spaced. Model (1) was fitted using the MIXED procedure in the SAS\textsuperscript{\textregistered} System (SAS 2003), Release 9.1.3.

We excluded water consumption measures that met either of the following two criteria from the model building: water consumption measurements of less than 7 g/day, which were considered unrealistic; and measures where laboratory notes indicated suspected slow water drip that, if included, would introduce bias and potentially affect model performance. Using the first criterion, we excluded 24 measurements (18 Gestation, 1 Lactation, 4 female, and 1 male Postweaning) from the model fitting; further analysis confirmed that excluding these measurements only negligibly affected the model estimates. Using the second criterion, we excluded another 83 measurements (39 Gestation, 44 Lactation, 0 female, and 0 male Postweaning).

We used a two-phased model building process to determine the appropriate set of covariates \( \{X_{t(ij)} \ldots X_{m(ij)}\} \) (for some number \( m \)) to include in Model (1) for predicting (ln-transformed) ADWC\textsubscript{T}. In the first phase, we assessed the predictive power of each candidate covariate individually and selected a preliminary set of covariates. In the second phase, we used a stepwise procedure to identify the set of candidate covariates that resulted in the best predicting model for ADWC\textsubscript{T}. This two-phased model building process is depicted in Figure 1.

---

![Figure 1. Model-building procedure. We built the regression models using a stepwise selection process. Phase 1: We initially selected candidate covariates through bivariate modeling and analyzed Pearson correlation coefficients. Phase 2: We specified a base model, which included the Time \( T(ij) \) variable and no covariates. Through subsequent steps, we built \( m \) models, with each step adding one of \( m \) candidate covariates to the model from the previous step. We added the candidate covariate yielding the best fit model to the model. The process iterated until we had included in the final model all of the candidate covariates meeting the statistical criteria. (*Note](image-url)
that we added quadratic terms to the list of candidate coefficients after the addition of linear terms to the model.

In the first phase, we individually evaluated each linear candidate covariate shown in Table 2 to assess the extent of its statistical association with (log-transformed) ADWC; in isolation, without confounding from other potential covariates. We evaluated the covariates by running a “bivariate” version of Model (1) for each candidate covariate, where the only two factors included in the model were the covariate and $\text{Time}_{t(ij)}$, the gestation day or postnatal day on which the end of the water consumption interval occurred. Within each bivariate model fit, we evaluated the effect of the covariate by testing the significance of the slope coefficient $\beta_1$ (F-test with $p$ value $[p] < 0.10$) and noting the goodness of fit across models using the value of the Akaike Information Criteria (AIC) statistic. (AIC is a goodness of fit measure useful for comparing statistical models that are fit to the same data set. The model with the lowest AIC value is considered the best model.) We also calculated Pearson correlation coefficient values (using the CORR procedure in SAS) to identify those pairs of candidate covariates that were highly correlated, thereby noting which covariates were likely to have a high degree of collinearity if we included them in the model at the same time.

We used the outcome of the first phase to inform the stepwise process used in the second phase, which yielded the final set of water consumption models. In the second phase, starting with only the $\text{Time}_{t(ij)}$ factor in the model, we used a forward selection procedure to add one candidate covariate at a time to Model (1), keeping sample sizes constant and assessing the improvement in fit with each addition through the likelihood ratio statistic. In a given step of this procedure, we inserted each of the remaining candidate covariates in isolation into the model from the previous step and refit the model (using PROC MIXED in SAS), noting the $p$ value from the likelihood ratio test and assuming a chi-square distribution for the test statistic. (Because there is no automated model selection function for PROC MIXED, we conducted the forward selection procedure manually.) We then repeated this process for each candidate covariate. If any of the $p$ values from these refits was $< 0.05$, we updated the model by adding the covariate associated with the smallest $p$ value (which was deemed to yield the greatest improvement in the model fit). We followed additional steps until no further covariates could be added.

We also applied a similar backward procedure to remove any covariate from the model that had a $p$ value $> 0.10$. This procedure helped to confirm that collinearity among the model covariates would be minimized. We exempted and did not remove covariates that corresponded to time, body weight at the end of the interval, and any linear terms in the presence of significant quadratic terms.

The set of candidate covariates we used in this model building exercise excluded certain covariates that were logically correlated with others (e.g., body weight at interval beginning and end). However, in general, we allowed the forward selection process to consider all factors and then confirmed that the final model did not include factors that were highly correlated.

We permitted two covariates — average body weight at end of interval and the time (i.e., day) covariate — to have both linear and quadratic terms in the model. We allowed quadratic terms for a curvilinear relationship within the ln-linear model and added them to the model only if the linear term was included as well. We did not consider quadratic terms for the other variables because they would have been difficult to justify from a practical standpoint.

Results

Model Building: First Phase

As described above, model development occurred in two phases. The results of the first phase are provided below (with significance noted at the 0.10 level). The list of candidate covariates for each model is provided in Table 2.

Gestation Model

All linear candidate covariates except sulfate concentration were significantly associated with water consumption. Increased ADWC; was also associated with higher body weight at end of interval, greater average daily body weight change (percent and absolute) during the interval, heavier total litter weight, longer intervals, number of live and dead pups found at birth, and higher sodium concentrations.

Lactation Model
All linear candidate covariates were significantly associated with ADWC. Increased ADWC was also associated with higher body weight at end of interval, greater average daily and percent average body weight change during the interval, number of live pups, heavier total litter weight, and higher sodium and sulfate concentrations. In addition, decreased ADWC was associated with longer intervals.

**Female Postweaning Model**

With the exception of length of water consumption interval, all linear candidate covariates were significantly associated with ADWC. Increased ADWC was associated with greater average daily body weight change (percent and absolute) during the interval, and higher average body weight at end of interval. Decreased ADWC was associated with increased number of rats in a cage.

**Male Postweaning Model**

With the exception of average percent daily weight gain, all linear candidate covariates were significantly associated with ADWC. Increased ADWC was associated with greater average daily body weight change during the interval, higher average body weight at end of interval, increased number of rats in a cage, and longer water consumption intervals. However, the consumption interval effect may not be of practical importance because the effect of interval length is relatively small, and the analysis data included intervals of 3 and 4 days only.

**Model Building: Second Phase**

The first phase provided necessary information for building the final water consumption models in the second phase, thereby allowing those covariates most strongly associated with ADWC to enter the models first. In Table 3, the outcomes of the stepwise selection process are presented, with separate columns for each of the four different models and each row specifying the estimate, standard error, and p value of the intercept \(\mu\) and slope coefficients \(\beta_1, \ldots, \beta_m\) for those covariates retained within the final model. Most of the slope estimates specified in Table 3 were statistically significant \((p < 0.05)\), with the exception of body weight at interval end for the female Postweaning model \((p = 0.38)\), which was forced in the model, along with its borderline significant quadratic term \((p = 0.10)\). The presence of a statistically significant quadratic term for a given covariate indicates that there is a curvilinear relationship between that covariate and ADWC.

Table 3  Estimated intercept and slope coefficients (with standard errors) for covariates in the final models for predicting log-transformed average daily water consumption interval

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Gestation model</th>
<th>Lactation model</th>
<th>Postweaning model (females)</th>
<th>Postweaning model (males)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept term overall mean water consumption</td>
<td>1.2941 (0.208)</td>
<td>2.6028 (0.148)</td>
<td>0.3773 (0.16)</td>
<td>1.8818 (0.171)</td>
</tr>
<tr>
<td>(p value)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Body weight (g) at end of water consumption interval</td>
<td>0.0115 (0.0013)</td>
<td>0.0026 (0.0004)</td>
<td>-0.0016 (0.0019)</td>
<td>0.0012 (0.0001)</td>
</tr>
<tr>
<td>(p value)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.380</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Variable</td>
<td>Quadratic body weight (g) at end of water consumption interval</td>
<td>Average daily change in body weight (g) during the interval</td>
<td>Percent average daily change in body weight during the interval</td>
<td>Gestation/postnatal day on interval end (Time&lt;sub&gt;ij&lt;/sub&gt;)</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>----------------------------------------------------------</td>
<td>----------------------------------------------------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>-0.00001 (0.0000017)</td>
<td>0.0066 (0.0018)</td>
<td>NA</td>
<td>-0.0187 (0.0028)</td>
</tr>
<tr>
<td></td>
<td>&lt;0.001</td>
<td>0.0066 (0.0018)</td>
<td>NA</td>
<td>0.036 (0.0016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0066 (0.0018)</td>
<td>NA</td>
<td>0.1217 (0.0104)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td>NA</td>
<td>0.0469 (0.0048)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td>NA</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td>NA</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td>NA</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td>NA</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td>NA</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Total litter weight recorded during initial/most recent litter examination | NA | NA
--- | --- | ---
Number of rats in a cage | -0.081 (0.025) | -0.06 (0.0175)
 | 0.002 | <0.001

\(^a\)Note that the impact of the slope coefficients on water consumption is a function of the slope coefficient times the value of its associated variable.

\(^b\)Quadratic body weight at interval end was removed from the list of potential covariates because its standard error was found to be inestimable in model-building.

\(^c\)Includes number found dead for the Gestation model.

\(^d\) NA, not applicable (term not included in the model).

Note that the final models in Table 3 do not contain all of the linear covariates found to be important when the covariates were assessed individually and in isolation (Table 2). In addition, they can include quadratic functions of certain covariates that were not assessed in the individual covariate analyses. This possibility is due to the effect of a covariate that may no longer be significant when in the presence of other covariates, and the model building process would thus exclude that covariate from the final model.

In the final Gestation and Lactation models (Table 3), the directional influence of each covariate was consistent with the results of the individual models (Table 2). However, in the Female Postweaning model, higher average body weight at end of interval was associated with greater and less ADWC\(_t\) in the individual and final models, respectively. This difference is likely offset by the inclusion of the quadratic term for this covariate in the final model, which is associated with greater ADWC\(_t\). Similarly, in the Male Postweaning model, a decreased number of rats in a cage was associated with greater and less ADWC\(_t\) in the individual and final models, respectively. The reason for this directional change in influence is unclear except to note that two additional terms that impact the ADWC\(_t\) — percent average daily change in body weight during the interval and the quadratic term for postnatal day at interval end — are present in the final model but were absent in the individual model.

The Male Postweaning model contains a positive coefficient for the average daily change in body weight and a negative coefficient for percent change in body weight. Clearly rats that gain more weight during an interval are likely to drink more water. In addition, if two rats gain the same amount of weight but only one had a higher percentage increase (equivalent to starting at a lower weight), then that rat might drink water at a different rate from the rat that started at a higher weight.

Our evaluation of consistency across models produced the following findings: (1) Only the variable Time\(_{ij}\) was highly significant \((p < 0.001)\) across all four models. (2) For both the female and male Postweaning models, the two variables that had the common characteristic of being highly significant were the quadratic term for Time\(_{ij}\) \((p < 0.001)\) and the number of rats in a cage \((p < 0.002)\). (3) For both the Lactation and Gestation models, the two variables that had the common characteristic of being highly significant were body weight at end of interval \((p < 0.001)\) and sodium concentration \((p < 0.006)\).

In Figure 2, we depict the goodness-of-fit (coefficient of determination; \(r^2\)) of each model in Table 3 relative to the data used in model development by plotting the model-predicted ADWC\(_t\) versus the consumption as measured within the studies involved in the development process. Gestation and Lactation models generally predicted ADWC\(_t\) well \((r^2 = 0.89 \text{ and } 0.92, \text{ respectively})\); however, their effectiveness was reduced for dams with higher consumption rates, as reflected by increased variability. The female and male Postweaning models also generally predicted ADWC\(_t\) well \((r^2 = 0.67 \text{ and } 0.79, \text{ respectively})\) but underestimated rats with very high water consumption rates.
Hypothetical Example of Model Application

In Table 4, we illustrate an application of the Lactation model to predict ADWC and approximate 95% prediction intervals for individual rats. In the hypothetical example, three lactating rats share a common water source. In Table 4, the values of the six Lactation model covariates are listed for each rat (as noted in Table 3).

Table 4. Covariate values and predicted water consumption rates for hypothetical example where three lactating rats share a common water source

<table>
<thead>
<tr>
<th>Lactation Model Covariates</th>
<th>Rat 1</th>
<th>Rat 2</th>
<th>Rat 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body weight (g) at end of water consumption interval</td>
<td>350</td>
<td>279</td>
<td>333</td>
</tr>
<tr>
<td>Average daily change in body weight (g) during the interval</td>
<td>0.25</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>Lactation day on interval end</td>
<td>13</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Water consumption interval length (no. of days)</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Sodium concentration in water (g/L)</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Number of live pups recorded during initial litter examination</td>
<td>8</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td><strong>Model Predictions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model Estimated Consumption (g/day)</td>
<td>78.65</td>
<td>64.46</td>
<td>80.24</td>
</tr>
</tbody>
</table>
Applying the Lactation model from Table 3, the predicted ln-transformed daily water consumption for Rat 1 is calculated by multiplying the value of the covariate provided in Table 4 by the value of the corresponding slope parameter taken from Table 3 as follows:

\[
\text{Natural Log ADWC}_1 = 2.6028 + 0.0026 \times 350 + 0.0066 \times 0.25 + 0.036 \times 13 - 0.015 \times 4 + 0.2389 \times 0.9 + 0.0284 \times 8 = 4.36
\]

Exponentiating this result to predict ADWC for Rat 1 is written as

\[
\text{ADWC}_1 = \exp(4.36) = 78.65 \text{ g/day.}
\]

Similarly, based on their covariate data, Rats 2 and 3 are predicted to consume 64.46 and 80.24 g/day, respectively. These predictions can then be used to distribute the actual total water consumed on a daily basis across the three rats via the common water source, resulting in a percentage of the total water source consumed by each rat. For Rat 1, this calculation would be

\[
\% \text{ Water Consumed Daily} = \frac{78.65}{(78.65 + 64.46 + 80.24)} = 35.2\%.
\]

If the water consumption for these three rats totaled 250 g/day, then the estimated water consumption, adjusted for total consumption, for Rat 1 is

\[
\text{Adjusted Estimated ADWC}_1 = 35.2\% \times 250 \text{ g} = 88.03 \text{ g.}
\]

Rats 2 and 3 are predicted to consume 72.15 and 89.81 g/day, respectively.

Because Model (1) includes mixed and nested effects, the calculation of standard errors associated with predicted values and, therefore, prediction intervals can be quite complex. By utilizing a direct approach, which integrated "dummy" records (only covariate values specified) with records containing the observed model data, we calculated a predicted standard error for each record employing the MIXED procedure in SAS®. We then used the standard error to calculate 95% prediction intervals with the following formula:

\[
\exp[\text{Predicted Consumption} - \ln(\text{Total Observed Consumption} / \text{Total Predicted}) \pm Z \times \text{Predicted Standard Error}],
\]

where \(Z=1.96\) is the standard normal deviate associated with the confidence coefficient of 95%.

The calculation above incorporates a finite population adjustment. Model diagnostics showed that the assumption of ln-linearity appeared to be valid. We employed SAS® to calculate standard errors for Rats 1 through 3, using the approach outlined above, with results provided in Table 4. For Rat 1, the 95% prediction interval would be

\[
\exp[4.36 - \ln((78.65 + 64.46 + 80.24)/250) \pm 1.96 \times 0.2196] = (57.3, 135.5).
\]

Discussion

We developed a set of linear models for predicting individual animal water consumption for gestating, lactating, and postweaning animals receiving water from a common source. Model applications are straightforward and predict both mean and 95% prediction intervals for ADWC\textsubscript{i} rates in individual rats. The \(R^2\) values for the model fits were good, ranging from 0.67 to 0.92. In general, the
ADWC predictions of the Gestation and Lactation models were consistent with measured data from comparable rats; the Postweaning models were less robust. It is conceivable that the multiple rats per cage in the postweaning data could have affected the male Postweaning model results. Because our model was based on averages, information on outliers within the cage was lost. Furthermore, the smaller sample sizes for the male and female postweaning rats would also limit model sensitivity.

The model covariates include attributes of animals that are routinely measured in rodent bioassays (e.g., age, body weight, litter size, duration of consumption interval). When measurement of individual animal water consumption rates is not possible or practical, scientists requiring more accurate estimates of daily water intake than provided with these models will need to obtain data that meet more stringent criteria on quality and quantity and that provide a wider range of possible predictor parameters than were available for model development. These investigators might need to consider more mathematically complex classes of models.

For certain studies used in developing the linear models, daily water consumption rates appeared to be higher than average, and the presence of irregularities in water delivery (e.g., leaks, blockages) can lead to increased uncertainty in consumption measurements. Such possibilities might be the reason for observing and classifying some low-end consumption values as statistical outliers. This potential occurs when the models report higher water consumption for those animals whose actual water consumption is very low compared with other animals on study. Exclusion of water consumption values < 7 g/day resulted in better fitting models, especially for gestating rats.

Although we minimized collinearity among similarly defined covariates (such as percent and absolute average weight) to the extent possible, the presence of collinearity can complicate the interpretation of the model estimates. We therefore include the caveat that estimates of model parameters associated with specific predictor variables (i.e., covariates) should not be interpreted in isolation but, instead, need to be considered jointly across all parameters in the model. Additional model forms may be considered that have a reduced tendency to flatten at the lower end of the distribution, which can lead to overestimation in this region. In turn, additional data would be useful for characterizing model behavior in these extreme regions.

Conclusion

We developed the water consumption modeling approach presented in this article to use in analyzing DBP exposures to individual animals in the EPA Four Lab Study. As for any model application, it is most appropriate to use these models to estimate water consumption in rats of the same strain/stock (Sprague-Dawley) and of comparable age, pregnancy status, and body weight. Applications of these models to other strains/stocks of rats, to rats of different body weights, or to rats of different ages than those used to develop the models will likely yield less accurate predictions of water consumption. Nevertheless, when appropriately applied, the models described herein may be useful to researchers using a reproductive rodent toxicity study design in which a common water source among animals is a feature of the study. In addition, these models can be generalized and rebuilt with appropriate data sets using the methodology shown in this article for use in other types of bioassays where water consumption rates are not measured but data on the explanatory variables are available.

Acknowledgments

The authors acknowledge the thoughtful reviews of Drs. Harlal Choudhury, Nina Wang, and Shahid Parvez. This work was performed by Battelle for the US Environmental Protection Agency (EPA) under Contract EP-C-05-030.

The manuscript has been reviewed and approved for publication in accordance with the peer and administrative review policies of the US EPA. Approval does not signify that the contents necessarily reflect the views and policies of the Agency nor does mention of trade names constitute endorsement or recommendation for use.

References


Sloan CS, Klinefelter GR, Goldman JM, Vick KD, Fail PA, Tyl RW. 2005. Delayed preputial separation (PPS) and SP22 measurement in rats administered bromochloroacetic acid (BCA) in drinking water. Toxicol Sci 84(1-S):112.