

# Assessing Effects of An Intervention on Bottle-Weaning and Reducing Daily Milk Intake from Bottles in Toddlers Using Two-Part Random Effects Models

Yungtai Lo<sup>1\*</sup>

<sup>1</sup> *Albert Einstein College of Medicine*

*Abstract:* Two-part random effects models have been used to fit semi-continuous longitudinal data where the response variable has a point mass at 0 and a continuous right-skewed distribution for positive values. We review methods proposed in the literature for analyzing data with excess zeros. A two-part logit-log-normal random effects model, a two-part logit-truncated normal random effects model, a two-part logit-gamma random effects model, and a two-part logit-skew normal random effects model were used to examine effects of a bottle-weaning intervention on reducing bottle use and daily milk intake from bottles in toddlers aged 11 to 13 months in a randomized controlled trial. We show in all four two-part models that the intervention promoted bottle-weaning and reduced daily milk intake from bottles in toddlers drinking from a bottle. We also show that there are no differences in model fit using either the logit link function or the probit link function for modeling the probability of bottle-weaning in all four models. Furthermore, prediction accuracy of the logit or probit link function is not sensitive to the distribution assumption on daily milk intake from bottles in toddlers not off bottles.

*Key words:* Two-part model, semi-continuous variable, truncated normal, gamma regression, skew normal, Pearson residual, receiver operating characteristic curve.

## 1. Introduction

In many applications, the response variable has a point mass at 0 and a continuous right-skewed distribution for nonzero values. For example, in a study of health care cost and utilization, many patients will have no health care utilization, resulting in no expenditure during the study. In a study of a bottle-weaning intervention for reducing bottle use and daily milk intake from bottles in 12-month-old toddlers, many toddlers will be off bottles at age of 12-15 months, resulting in no daily milk intake from bottles during the study. We refer the response variable that can take the value of 0 or any value between its lower bound and upper bound as the semi-continuous variable. Since the semi-continuous variable has a continuous distribution except for a point mass at 0, estimates and inferences from regression models without taking this feature

---

\* Corresponding author.

into account can be seriously biased. Several approaches have been proposed in the literature to address semi-continuous data.

One approach is to collapse the semi-continuous response variable into categories and then apply logistic or probit models for subsequent analysis (Saei, Ward and McGilchrist, 1996). Whereas this approach is simple, information may lose through grouping the response outcomes into categories. There may be no sensible way to collapse the positive scale into categories to reflect the qualitative distinction between zero and nonzero responses.

An alternative is to use the Tobit model proposed by Tobin (1958) for analyzing household expenditures on durable goods in which data on demand for durable goods had values clustered at a limiting value, zero. Tobit models assume that a zero response arises from the underlying normal random variable that is censored at 0 by a random mechanism. Under this assumption, censored linear regression models can then be applied to model semi-continuous data. It is well known that Tobit models are sensitive to minor departures from normality. When the response variable is not censored normal, Poirier (1978) suggested using the Box-Cox power transformation to correct departures from normality. Powell (1986) proposed a semi-parametric least squares estimation based on symmetric censoring or truncation of the upper tail of the distribution of the response variable for Tobit models.

When the zeros are valid observed responses rather than censored values, Tobit models that assume a common stochastic process determining both the occurrence of a zero response and the intensity of a nonzero response cannot capture the underlying distribution of nonzero values because the meaning of the underlying normality becomes dubious and the zeros may respond differently to parameters or covariates (Duan *et al.*, 1983). Cragg (1971) extended the Tobit model by allowing the determination of the magnitude of a positive response to depend on different parameters or covariates from those determining the occurrence of a zero response. This model is referred to as the two-part model because the zeros and nonzeros are modeled separately in the literature. Specifically, the first part (Part 1) of such a model deals with the point mass at 0. The response in Part 1 is dichotomous (zero or nonzero). Logistic or probit regression is typically used for modeling the probability of observing a zero. The second part (Part 2) of the model deals only with positive values conditional on a nonzero response. Cragg (1971) suggested employing truncated normal or log-normal to model the mean response among positive values in Part 2. Blough, Madden and Hornbrook (1999) proposed using generalized linear models to model the nonzero responses. They suggested a quasi-likelihood based method for choosing link and variance functions. In practice, two-part models are specified using the same set of covariates in both parts. However, distinct sets of covariates may be used in two-part model applications.

Olsen and Schafer (2001) extended two-part models to repeated measures semi-continuous data. Like mixed effects models, they incorporated subject specific random intercepts into both the logistic model for modeling the probability of a nonzero response in Part 1 and the conditional linear regression model for modeling the mean response in Part 2 to account for repeated measures taken from the same subjects over time. The random effects from the two parts are assumed to be jointly normal and possibly correlated. If the covariance between the two random effects is zero, then the logistic regression in Part 1 and the conditional linear regression in Part 2 can be modeled separately as in the case of cross-sectional data. Assumptions of independent

random effects are often made in practice for computational feasibility. Su *et al.* (2009) showed that bias can be induced for regression coefficients in the second part of the model when random effects are truly correlated but misspecified as independent in two-part random effects models. See Lee *et al.* (2010), Liu *et al.* (2010), Mahmud *et al.* (2010), Lee and Xiang (2011), Li *et al.* (2011), and Mills (2013) for recent work on two-part modeling of semi-continuous data.

When a zero response can either be a valid observed response called a true zero or a partial observation censored at 0 as in Tobit analysis from a continuous distribution for the nonzero values, the logit or probit link function in Part 1 of two-part models can-not differentiate the true zero from the observation censored at 0. Moulton and Halsey (1995) proposed a zero-inflated Tobit model by allowing the possibility that some pro-portion of zeros come from a point mass distribution at 0 whereas the remaining zeros come from a censored log-normal distribution. Berk and Lachenbruch (2002) applied the mixture of a point mass at 0 and a Tobit model proposed by Moulton and Halsey (1995) to analyze longitudinal zero-inflated continuous data where a zero observation can either be a true zero or a left-censored value from a log-normal distribution. The probability of a true zero was modeled with logistic regression with a random intercept and the other values were modeled with censored log-normal regression with a random intercept.

In this paper, we wish to examine effects of a bottle-weaning intervention on reducing bottle use and daily milk intake from bottles over 12 months of follow up in toddlers aged 11 to 13 months. Brotanek *et al.* (2005) reported that children with prolonged bottle-feeding are at higher risk for iron deficiency. Sutcliffe *et al.* (2006) showed that children drinking milk from a bottle during the daytime are more likely to be iron depleted compared with cup-fed children. Excess protein intake from milk and formula can increase the risk of overweight or obesity. Several studies have shown that prolonged bottle use is associated with overweight or obesity (Bonuck and Kahn, 2002 ; Gooze *et al.*, 2011). The American Academy of Pediatrics Committee on Nutrition (1992) recommended that infants be introduced to the cup around 6 months of age and be fully weaned from the bottle by 15 months to promote feeding skills. Toddlers off bottles will have daily milk intake from cups instead of bottles during the study. The repeated measures outcome variable, total daily milk intake from bottles, will take either a value of zero, indicating bottle-weaning, or a positive value measuring the amount of milk intake from bottles. As many toddlers will be weaned from a bottle by 15 months, the outcome variable will have a substantial proportion of zero values. We cannot use a Tobit model to examine effects of the bottle-weaning intervention on reducing the probability of drinking from a bottle and the amount of daily milk intake from bottles because it will require the assumption that an observation that has no daily milk intake from bottles is a censored zero from a toddler who drinks from a bottle but happens to not drink from a bottle in the current period, rather than a valid zero from a toddler who is weaned from bottles. This assumption implies that no toddlers are truly weaned from a bottle during the study. It is commonly perceived that weaned toddlers will not return to bottle use and hence, an observation that has no daily milk intake from bottles can be considered as a true zero. A zero-inflated Tobit model proposed by Moulton and Halsey (1995) may not be suitable for analyzing the total daily milk intake from bottles data because it will require the assumption that some of weaned toddlers will return to bottle use during the study period and hence, an

observation that has no daily milk intake from bottles is either a true zero from a toddler off bottles or a censored zero from a toddler who drinks from a bottle but happens to not drink from a bottle in the current period. Unlike Tobit or zero-inflated Tobit models, two-part models treat an observation that has no daily milk intake from bottles as a true zero from a toddler weaned from bottles and do not assume that some of weaned toddlers will return to bottle use during the study period. We propose the use of a two-part random effects model to model the response variable, total daily milk intake from bottles, that can take a value either from a point mass at 0 or from a right-skewed, continuous distribution and to assess effects of the bottle-weaning intervention on reducing the probability of bottle use and the intensity of total daily milk intake from bottles over 12 months of follow up in toddlers aged 11 to 13 months.

The remaining of this paper is organized as follows. Section 2 presents two-part random effects models with logistic or probit regression for modeling the probability of a nonzero response in Part 1 and with log-normal, truncated normal, gamma, or skew normal regression for modeling the intensity of positive values in Part 2. The four two-part random effects models given in Section 2 are applied to analyze semi-continuous data on total daily milk intake from bottles in Section 3. Section 3 also presents the receiver operating characteristic (ROC) curve for assessing model fit in Part 1 and the normal quantile-quantile (Q-Q) plot of Pearson residuals for assessing model fit in Part 2. In addition, results of the four two-part random effects models with the logit link function replaced by the probit link function in Part 1 are presented in Section 3. Discussions and concluding remarks are presented in Section 4.

## 2. Two-Part Random Effects Model for Repeated Measures Semi-Continuous Data

Let  $Y_{ij}$  denote a semi-continuous response variable for subject  $i = 1, \dots, m$  at occasion  $j = 1, \dots, n_i$ . The probability density function of  $Y_{ij}$  is given by

$$f(y_{ij}) \begin{cases} P(Y_{ij} = 0), & \text{if } y_{ij} = 0; \\ P(Y_{ij} > 0)f(y_{ij}|y_{ij} > 0), & \text{if } y_{ij} > 0; \\ 0, & \text{if } y_{ij} < 0. \end{cases}$$

The first part (Part 1) of the two-part model is to model the probability of a nonzero response. Conditional on a nonzero response, the second part (Part 2) of the model then models the intensity or level of positive values (Cragg, 1971; Olsen and Schafer, 2001).

### 2.1. Part 1 of the Two-Part Random Effects Model

Let  $P(Y_{ij} > 0) = p_{ij}$  and  $P(Y_{ij} = 0) = 1 - p_{ij}$ . A logistic or probit regression model with a random effect can be employed to predict which responses will be positive and which responses will be zero. Let  $u_{1i}$  be the random effect due to the  $i$ th subject affecting the probability  $p_{ij}$ , and it is assumed to follow a normal distribution with mean 0 and variance  $\tau^2$ . Conditional on  $u_{1i}$ , the logit link function for  $p_{ij}$  is given as

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = X_{1ij}^T \beta_1 + u_{1i}, \quad (1)$$

where  $X_{1ij}$  is a vector of covariates associated with the occurrence of a positive response for subject  $i$  at occasion  $j$ , and  $\beta_1$  is a vector of unknown regression parameters. Similarly, conditional on  $u_{1i}$ , the probit link function for  $p_{ij}$  is given as

$$p_{ij} = \Phi(X_{1ij}^t \beta_1 + u_{1i}), \quad (2)$$

where  $\Phi(\cdot)$  is the standard normal cumulative distribution function.

## 2.2. Part 2 of the Two-Part Random Effects Model

Conditional on positive values, several possible models for modeling the mean response of  $Y_{ij}$  can be applied, depending on the distribution assumption on  $Y_{ij}$ . We present four such models assuming that the positive values in Part 2 follow a log-normal distribution (Olsen and Schafer, 2001; Tooze *et al.*, 2002), a truncated normal distribution (Cragg, 1971), a gamma distribution (Blough *et al.*, 1999; Lee *et al.*, 2010), and a skew normal distribution (Chai and Baily, 2008; Mahmud *et al.*, 2010). Let  $u_{2i}$  be the unobserved random effect due to the  $i$ th subject affecting the intensity or level of positive values, and it is assumed to follow a normal distribution with mean 0 and variance  $v^2$ .

### 2.2.1. Log-Normal Distribution

If  $Y_{ij}$  conditional on  $Y_{ij} > 0$  and  $u_{2i}$  is assumed to follow a log-normal distribution with parameters  $\mu_{ij}$  and  $\sigma$ , then

$$f(y_{ij}|y_{ij} > 0, u_{2i}) = \frac{1}{\sigma y_{ij} \sqrt{2\pi}} \exp\left[-\frac{(\log y_{ij} - \mu_{ij})^2}{2\sigma^2}\right], \quad (3)$$

where  $\mu_{ij} = X_{2ij}^T \beta_2 + u_{2i}$ ,  $X_{2ij}$  is a vector of covariates associated with the intensity or level of a positive response for subject  $i$  at occasion  $j$ , and  $\beta_2$  is a vector of unknown regression coefficients. Conditional on  $Y_{ij} > 0$  and  $u_{2i}$ , the mean and variance of  $Y_{ij}$  are given respectively as

$$E(Y_{ij}|Y_{ij} > 0, u_{2i}) = e^{\mu_{ij} + (\sigma^2/2)}, \quad (4)$$

and

$$\text{Var}(Y_{ij}|Y_{ij} > 0, u_{2i}) = e^{2(\mu_{ij} + \sigma^2)} - e^{2\mu_{ij} + \sigma^2}. \quad (5)$$

### 2.2.2. Truncated Normal Distribution

If  $Y_{ij}$  conditional on  $Y_{ij} > 0$  and  $u_{2i}$  is assumed to follow a truncated normal distribution with parameters  $\mu_{ij}$  and  $\sigma$  (Barr and Sherrill, 1999), then

$$f(y_{ij} | y_{ij} > 0, u_{2i}) = \frac{\frac{1}{\sigma} \phi\left(\frac{y_{ij} - \mu_{ij}}{\sigma}\right)}{1 - \Phi\left(\frac{-\mu_{ij}}{\sigma}\right)} \quad (6)$$

where  $\phi(\cdot)$  is the standard normal density function,  $\Phi(\cdot)$  is the standard normal cumulative distribution function,  $\mu_{ij} = X_{2ij}^T \beta_2 + u_{2i}$ ,  $X_{2ij}$  is a vector of covariates associated with the intensity or level of a positive response for subject  $i$  at occasion  $j$ , and  $\beta_2$  is a vector of unknown regression coefficients. Conditional on  $Y_{ij} > 0$  and  $u_{2i}$ , the mean and variance of  $Y_{ij}$  are given respectively as

$$E(Y_{ij} | Y_{ij} > 0, u_{2i}) = \mu_{ij} + \sigma \frac{\phi\left(\frac{-\mu_{ij}}{\sigma}\right)}{1 - \Phi\left(\frac{-\mu_{ij}}{\sigma}\right)} \quad (7)$$

and

$$Var(Y_{ij} | Y_{ij} > 0, u_{2i}) = \sigma^2 \left[ 1 + \frac{\frac{-\mu_{ij}}{\sigma} \phi\left(\frac{-\mu_{ij}}{\sigma}\right)}{1 - \Phi\left(\frac{-\mu_{ij}}{\sigma}\right)} - \left( \frac{\phi\left(\frac{-\mu_{ij}}{\sigma}\right)}{1 - \Phi\left(\frac{-\mu_{ij}}{\sigma}\right)} \right)^2 \right]. \quad (8)$$

### 2.2.3. Gamma Distribution

If  $Y_{ij}$  conditional on  $Y_{ij} > 0$  and  $u_{2i}$  is assumed to follow a gamma distribution with parameters  $\alpha$  and  $\lambda_{ij}$ , then

$$f(y_{ij} | y_{ij} > 0, u_{2i}) = \frac{1}{\alpha^{\lambda_{ij}} \Gamma(\lambda_{ij})} y_{ij}^{\lambda_{ij}-1} e^{-y_{ij}/\alpha} \quad (9)$$

where  $\lambda_{ij}$  is the unknown shape parameter, and  $\alpha$  is the unknown scale parameter and is assumed constant. Conditional on  $Y_{ij} > 0$  and  $u_{2i}$ , the mean and variance of  $Y_{ij}$  are given respectively as

$$\begin{aligned} E(Y_{ij} | Y_{ij} > 0, u_{2i}) &= \alpha \lambda_{ij} \\ &= e^{X_{2ij}^T \beta_2 + u_{2i}} \end{aligned} \quad (10)$$

and

$$\begin{aligned} Var(Y_{ij} | Y_{ij} > 0, u_{2i}) &= \alpha^2 \lambda_{ij} \\ &= \alpha E(Y_{ij} | Y_{ij} > 0, u_{2i}) \end{aligned} \quad (11)$$

where  $X_{2ij}$  is a vector of covariates associated with the intensity or level of a positive response for subject  $i$  at occasion  $j$  and  $\beta_2$  is a vector of unknown regression coefficients.

### 2.2.4. Skew Normal Distribution

If  $Y_{ij}$  conditional on  $Y_{ij} > 0$  and  $u_{2i}$  is assumed to follow a skew normal distribution with parameters  $\mu_{ij}$ ,  $\sigma$ , and  $\delta$  (Chai and Bailey, 2008), then

$$f(y_{ij} | y_{ij} > 0, u_{2i}) = \frac{2}{\sqrt{\sigma^2 + \delta^2}} \phi\left(\frac{y_{ij} - \mu_{ij}}{\sqrt{\sigma^2 + \delta^2}}\right) \Phi\left(\frac{\delta}{\sigma} \frac{y_{ij} - \mu_{ij}}{\sqrt{\sigma^2 + \delta^2}}\right) \quad (12)$$

where  $\phi(\cdot)$  is the standard normal density function,  $\Phi(\cdot)$  is the standard normal cumulative distribution function,  $\mu_{ij} = X_{2ij}^T \beta_2 + u_{2i}$ ,  $X_{2ij}$  is a vector of covariates associated with the intensity or level of a positive response for subject  $i$  at occasion  $j$ ,  $\beta_2$  is a vector of unknown regression coefficients, and  $\delta$  is the skewness parameter. The distribution is skewed to right when  $\delta > 0$ , skewed to left when  $\delta < 0$ , reduces to the normal distribution when  $\delta = 0$ , and converges to the half normal distribution when  $\delta \rightarrow \infty$ . Conditional on  $Y_{ij} > 0$  and  $u_{2i}$ , the mean and variance of  $Y_{ij}$  are given respectively as

$$E(Y_{ij} | Y_{ij} > 0, u_{2i}) = \mu_{ij} + \delta \sqrt{\frac{2}{\pi}} \quad (13)$$

and

$$Var(Y_{ij} | Y_{ij} > 0, u_{2i}) = \sigma^2 + \delta^2 \left(1 - \frac{2}{\pi}\right). \quad (14)$$

### 2.3. Maximum Marginal Likelihood Estimation

Part 1 and Part 2 are linked through correlated random effects. Assume that the random effects,  $u_i = (u_{1i} \ u_{2i})^T$ , follow a bivariate normal distribution with mean  $(0 \ 0)^T$  and variance-covariance matrix  $\Sigma$  given as

$$\begin{pmatrix} \tau^2 & \rho\tau\nu \\ \rho\tau\nu & \nu^2 \end{pmatrix} \quad (15)$$

where  $\rho$  is the correlation of  $u_{1i}$  and  $u_{2i}$ . The correlation coefficient  $\rho$  quantifies the cross-part correlation between Part 1 for determining the occurrence of a nonzero response and Part 2 for determining the level of a positive response. Let  $z_{ij}$  be 1 if  $y_{ij} = 0$  and 0 if  $y_{ij} > 0$ . Given the random effects, the marginal likelihood contribution from subject  $i$  is given as

$$L_i(\beta_1, \beta_2, |\Sigma) = \int \int \prod_{j=1}^{m_i} (1 - p_{ij})^{z_{ij}} [p_{ij} f(y_{ij} | y_{ij} > 0, u_{2i})]^{1-z_{ij}} g(u_{1i}, u_{2i}) du_{1i} du_{2i} \quad (16)$$

where  $m_i$  is the number of repeated measurements on subject  $i$ ,  $p_{ij}$  is given in (1) if the logit link function is used in Part 1 or given in (2) if the probit link function is used in Part 1,  $f(y_{ij} | y_{ij} > 0, u_{2i})$  is given in (3), (6), (9), or (12), depending on the distribution assumption on  $y_i > 0$ , and  $g(u_{1i}, u_{2i})$  is the bivariate normal density of the two random effects  $u_{1i}$  and  $u_{2i}$ . Let  $n$  be the total number of subjects. Assume that the measurements on different subjects are independent. The likelihood to be maximized is

$$L(\beta_1, \beta_2, |\Sigma) = \prod_{i=1}^n L_i(\beta_1, \beta_2, \Sigma). \quad (17)$$

The likelihood in (17) involves integrating a nonlinear function over the two random effects given in (16). Numerical methods incorporated with some technique to approximate the integral are needed to find the maximum likelihood estimators for  $\beta_1$ ,  $\beta_2$ , and  $\Sigma$ . Olsen and Schafer (2001) used a sixth-order Laplace approximation to first approximate the marginal likelihood and

then used an approximate Fisher scoring algorithm to maximize the likelihood. Tooze *et al.* (2002) used the quasi-Newton algorithm along with adaptive Gaussian quadrature to maximize the likelihood. This approach can be easily implemented in standard statistical software such as SAS, Stata, or R to find the maximum likelihood estimates.

### 3. An Example

Consider data from  $n=245$  parent-child dyads from a bottle-weaning intervention study. Parents of toddlers aged 11 to 13 months who were consuming at least 2 bottles of milk per day were randomized to the bottle-weaning intervention or control group. Nutritionists at Women, Infants, and Children (WIC) Supplemental Feeding Program sites delivered the educational intervention, counseling guided by a flip-chart. The flip-chart presented message about healthy weight, dental caries, and iron deficiency anemia effects from bottle-weaning. The intervention group also received a pamphlet to share with family members and a lidded, two-handled 6 ounce sippy cup with a hard spout and no internal leak proof valve. The intervention recommended that parents gradually replace bottles with cups. Details on the educational intervention for bottle-weaning are given in Hyden, Kahn and Bonuck (2012). Bottle use and total daily milk intake in ounces from bottles were assessed at baseline, 15, 18, 21, and 24 months old. Details on the study design are given in Bonuck *et al.* (2010). We wish to assess effects of the intervention on bottle-weaning and reducing daily milk intake from bottles over 12-months of follow-up.

Of 245 toddlers aged 11 to 13 months who were consuming at least 2 bottles of milk per day, 118 toddlers were randomly assigned to the intervention and 127 toddlers were randomly assigned to the control. There were no significant differences in baseline child characteristics such as age, sex, total sleep per day, and race, number of bottles at baseline, total daily milk intake from bottles at baseline, total daily energy intake at baseline, and weight for length at baseline between the intervention and control. There were no significant differences in caregiver characteristics such as age, education, and child care between the intervention and control. Fifty-eight children had one repeated measure, 64 children had two repeated measures, 65 children had three repeated measures, and 58 children had 4 repeated measures post intervention. The histogram of total daily milk intake in ounces from bottles is given in Figure 1.

Figure 1 shows that the response variable, total daily milk intake from bottles, was a semi-continuous variable with 246 (40.1%) values equal to zero and a continuous right-skewed distribution among the positive values. The zeros were from 126 children (67 intervention and 59 control) who were weaned from a bottle and corresponded to no daily milk intake from bottles at follow-up. None of these 126 weaned children returned to bottle use during 12 months of follow-up. Hence, the observations that had no daily milk intake from bottles are considered as true zeros, not as censored zeros that would have arisen from toddlers who drink from a bottle but happen to not drink from a bottle in the current period. Two-part random effects models given

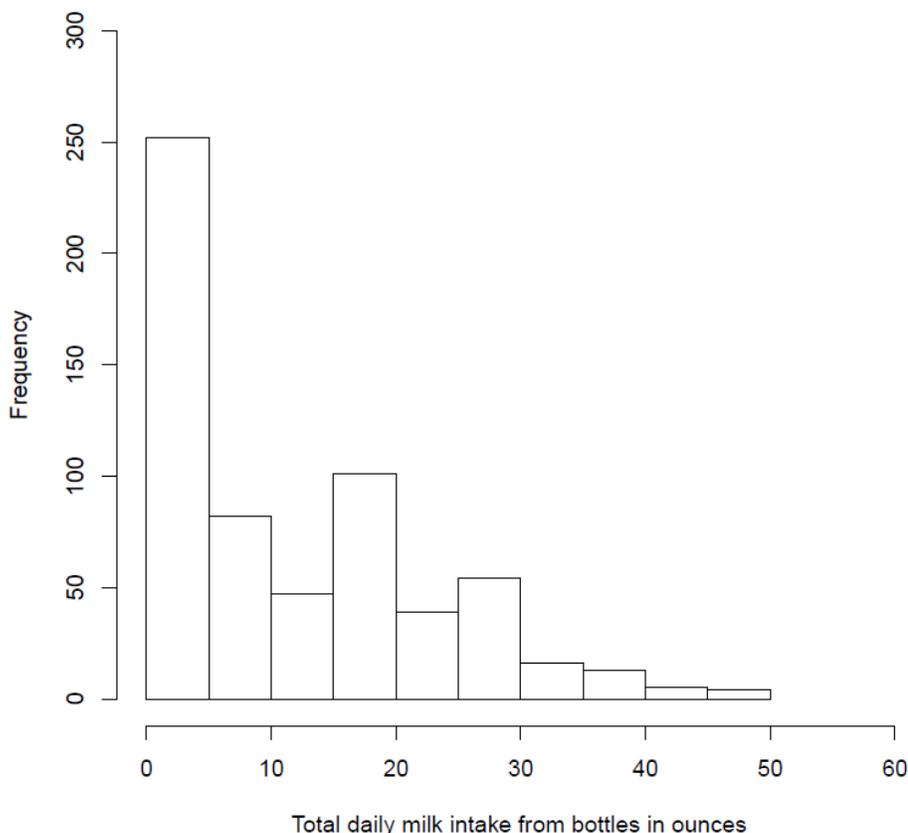


Figure 1: Histogram of total daily milk intake from bottles from 245 children aged 11 to 13 months

in Section 2 can be used to assess effects of the intervention on bottle-weaning and reducing total daily milk intake from bottles accounting for the excess zeros and repeated measurements on each subject over 12 months of follow-up. Part 1 of such a model models the probability that a toddler is weaned from a bottle or drinks from a bottle. Conditional on drinking from a bottle, Part 2 of the model then models how much milk a toddler drinks.

We first fitted data on total daily milk intake from bottles to two-part random effects models with the logit link function given in (1) for modeling the probability of drinking from a bottle in Part 1 and log-normal regression given in (3) denoted by Model 1, truncated normal regression given in (6) denoted by Model 2, gamma regression given in (9) denoted by Model 3, or skew normal regression given in (12) denoted by Model 4 for modeling the intensity of total daily milk intake from bottles conditional on drinking from a bottle in Part 2, depending on the distribution assumed for the positive values. We began our modeling including months post baseline, intervention, total daily milk intake from bottles at baseline, an interaction between intervention and baseline daily milk intake from bottles, and an interaction between intervention and months post baseline in both Part 1 and Part 2 for all four models. Starting values for the parameters in Part 1 can be obtained from fitting a logistic regression model with a subject-specific random

intercept using the binary response only. Similarly, starting values for the parameters in Part 2 can be obtained from fitting a mixed effects model limited to the positive values. The SAS NLMIXED procedure with the adaptive Gaussian quadrature and quasi-Newton method was used to obtain the maximum likelihood estimates for the parameters in both Part 1 and Part 2.

The parameter estimates and their corresponding standard errors for the four two-part models are given in Table 1. The logit link for modeling the probability of drinking from a bottle included terms for total daily milk intake from bottles at baseline, intervention, and months post baseline in each of four models. The four models gave nearly identical estimates for the parameters and random intercept variance  $\tau^2$  in Part 1. Specifically, higher baseline daily milk intake from bottles was associated with higher odds of drinking from a bottle at follow-up. The intervention

Table 1: Parameter estimates and standard errors for the four two-part random effects models fitted to the total daily milk intake from bottles data

Variable	Model 1 <sup>a</sup>	Model 2 <sup>a</sup>	Model 3 <sup>a</sup>	Model 4 <sup>a</sup>
Part 1: drinking from a bottle				
Intercept	4.24(0.62)***	4.23(0.62)***	4.23(0.62)***	4.28(0.62)***
Baseline daily milk intake from bottles <sup>b</sup>	0.06(0.03)*	0.07(0.03)*	0.07(0.03)*	0.07(0.03)*
Intervention	-1.34(0.52)*	-1.32(0.52)*	-1.33(0.52)*	-1.38(0.53)**
Months post baseline	-0.38(0.06)***	-0.38(0.06)***	-0.38(0.06)***	-0.38(0.06)***
Part 2: modeling nonzero data				
Intercept	2.94(0.07)***	21.59(1.22)***	3.06(0.07)***	11.52(1.53)***
Baseline daily milk intake from bottles <sup>b</sup>	0.01(0.003)**	0.23(0.06)**	0.01(0.003)**	0.16(0.05)**
Intervention	-0.20(0.07)**	-3.86(1.29)**	-0.19(0.06)**	-3.17(0.98)**
Months post baseline	-0.03(0.01)**	-0.69(0.16)***	-0.03(0.01)**	-0.49(0.15)**
$\alpha$	-	-	5.29(0.53)***	-
$\delta$	-	-	-	12.11(1.19)***
$\sigma$	0.45(0.02)***	7.87(0.47)***	-	3.35(0.90)**
$\tau^2$	9.27(2.70)**	9.26(2.70)**	9.28(2.70)**	9.50(2.77)**
$\nu^2$	0.09(0.03)**	34.11(9.22)**	0.07(0.02)**	14.70(5.99)*
$\rho$	0.59(0.16)**	0.62(0.17)**	0.64(0.17)**	0.74(0.21)**
log-likelihood	-1615.0	-1612.6	-1603.4	-1613.9
AIC	3253.9	3249.2	3230.8	3253.8
BIC	3295.9	3291.2	3272.8	3299.3

<sup>a</sup>Model 1: a two-part logit-log-normal random effects model; Model 2: a two-part logit-truncated normal random effects model; Model 3: a two-part logit-gamma random effects model; Model 4: a two-part logit-skew normal random effects model.

<sup>b</sup>Baseline total daily milk intake from bottles was centered at median.

\*p-value < 0.05; \*\*p-value < 0.01; \*\*\*p-value < 0.0001.

reduced odds of drinking from a bottle compared to the control. Each month post baseline was associated with decreased odds of drinking from a bottle at follow-up. There were no significant

inter-action between intervention and baseline daily milk intake from bottles, and interaction between intervention and months post baseline in any of the four models. The estimate of the random intercept variance  $\tau^2$  was about 3.4 times larger than its standard error in all four models, which suggests that there was considerable subject-specific variability in trajectories of drinking from a bottle over time. Conditional on drinking from a bottle, Part 2 for modeling the intensity of total daily milk intake from bottles included terms for baseline total daily milk intake from bottles, intervention, and months post baseline in each of four models. The parameter estimates for the intercept and the three factors in Model 1 and Model 3 are similar whereas the parameter estimates in Model 2 and Model 4 are similar. In all four models, higher baseline daily milk intake predicted higher total daily milk intake from bottles at follow-up conditional on drinking from a bottle. Conditional on drinking from a bottle, toddlers in the intervention had less daily milk intake from bottles compared to toddlers in the control. Conditional on drinking from a bottle, each month post baseline was associated with reduced daily milk intake from bottles. There were no significant interaction between intervention and baseline daily milk intake from bottles, and interaction between intervention and months post baseline in any of the four models. The estimate of the random intercept variance  $v^2$  was statistically significant in all four models. Model 2 and Model 4 suggest considerable amounts of subject-specific variability in trajectories of total daily milk intake from bottles over time whereas Model 1 and Model 3 suggest non-trivial amounts of subject-specific variance. The estimated correlation coefficient between the two random intercepts was 0.59 ( $p = 0.0004$ ) in Model 1, 0.62 ( $p = 0.0003$ ) in Model 2, 0.64 ( $p = 0.0002$ ) in Model 3, and 0.74 ( $p = 0.0005$ ) in Model 4. This substantially positive cross-part correlation suggests the higher the odds of drinking from a bottle, the greater the amount of daily milk intake from bottles. That is, if a toddler is more likely to drink from a bottle over time, the toddler is more likely to have large amount of daily milk intake from bottles.

We assessed model fit in Part 1 through the receiver operating characteristic (ROC) curve. The ROC curve displays the relationship between sensitivity and  $1 - \text{specificity}$  across all possible cut-points for defining the positivity (bottle use). The area under the ROC curve measures the ability of the logit model in Part 1 to discriminate between those toddlers who drink from a bottle and those who are weaned from a bottle. The higher area under the ROC curve, the better the overall performance of the logit model. We first estimated the predicted probabilities of drinking from a bottle from the logit link with a random intercept in Part 1 and then used these predicted probabilities to construct the ROC curve. We defined sensitivity as the proportion of true positives (bottle use) correctly classified by the model and specificity as the proportion of true negatives (bottle-weaning) correctly classified by the model. For example, if the predicted probability exceeds 0.5, we will classify a toddler as drinking from a bottle. Using the cut-off point 0.5, the logit part in all four models yielded a sensitivity of 95.6% and a specificity of 91.1%. The ROC curves for Model 1, Model 2, Model 3, and Model 4 are given in Figure 2. Figure 2 indicates that the four models had almost identical ROC curves. All the curves followed closely the left-hand border and then the top border of the ROC space, indicating that the logit model in Part 1 gave excellent prediction of toddlers drinking from a bottle in all four models. The estimated area under the ROC curve was 0.9795 for Model 1, 0.9794 for Model 2,

0.9795 for Model 3, and 0.9796 for Model 4, with the corresponding standard errors given by 0.00465, 0.00467, 0.00468, and 0.00462. The estimated areas under the ROC curves in four models were nearly identical. An area of 0.9795 can be interpreted as follows: if we select two toddlers at random, one drinking from a bottle and one weaned from a bottle, the probability is 0.9795 that a toddler drinking from a bottle will be classified as bottle use by the model. We conclude from the above results that model fit in Part 1 is not sensitive to the distribution assumption on the response variable in Part 2.

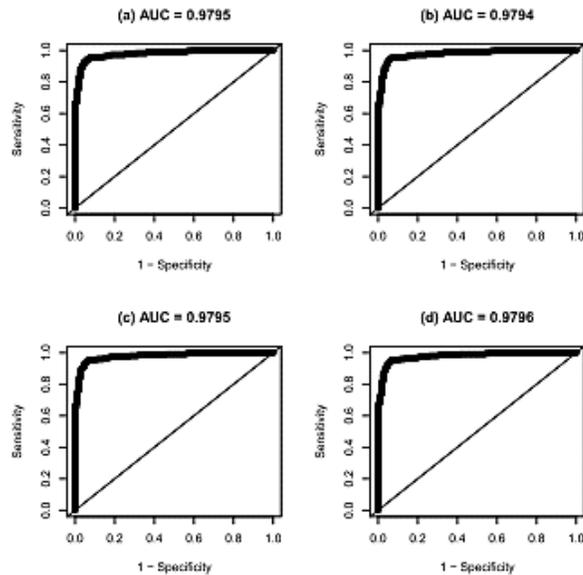


Figure 2: The receiver operating characteristic curves for the logit link function in Part 1 for (a) Model 1: a two-part logit-log-normal random effects model, (b) Model 2: a two-part logit-truncated normal random effects model, (c) Model 3: a two-part logit-gamma random effects model, and (d) Model 4: a two-part logit-skew normal random effects model. AUC stands for the area under the curve.

Model fit in Part 2 conditional on drinking from a bottle was assessed through the quantile-quantile (Q-Q) plot of Pearson (or standardized) residuals. Pearson residuals in Part 2 were calculated by taking the difference between actual and fitted values and dividing by the estimated standard deviation. The Fitted value and estimated standard deviation in Part 2 were obtained respectively from (4) and (5) for Model 1, (7) and (8) for Model 2, (10) and (11) for Model 3, and (13) and (14) for Model 4. Under a well-specified model in Part 2, Pearson residuals are expected to be standard normal when the sample size is large. The Q-Q plots of Pearson residuals against the standard normal distribution are given in Figure 3. There are departures from the line in Figure 3(a) for Model 1. However, these do not suggest severe divergence from the standard normal distribution. The quantiles of the theoretical and data distribution agree except in the lower tail of the plots in Figure 3(b) for Model 2 and Figure 3(c) for Model 3. Although not all

plotted points fall on or near the line, Figure 3(b) and Figure 3(c) do not indicate substantial departures from the standard normal distribution. There are departures from the line in the lower

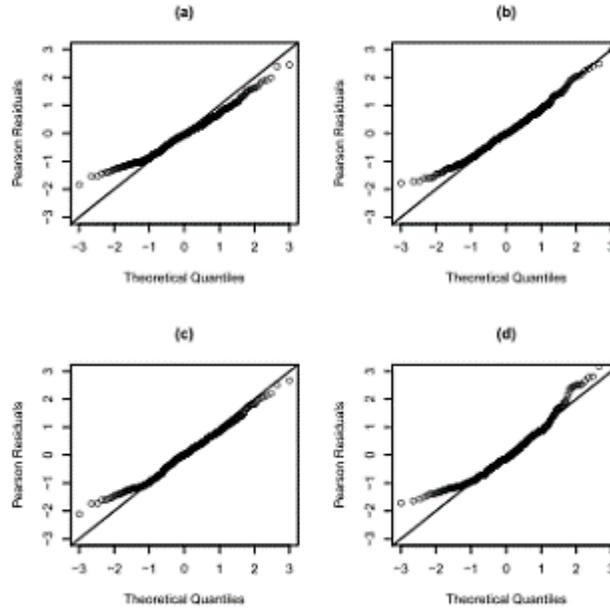


Figure 3: The Q-Q plots of Pearson residuals against the standard normal distribution in Part 2 for (a) Model 1: a two-part logit-log-normal random effects model, (b) Model 2: a two-part logit-truncated normal random effects model, (c) Model 3: a two-part logit-gamma random effects model, and (d) Model 4: a two-part logit-skew normal random effects model.

and upper tails of the Q-Q plot in Figure 3(d) for Model 4, indicating a poor fit in Part 2. In summary, the lower tail is not well captured by any of four models. Except in the lower tail, the Q-Q plots for Model 2 and Model 3 are close to the 45-degree straight line, suggesting Model 2 and Model 3 fitted to the data better than Model 1 and Model 4 conditional on drinking from a bottle in Part 2.

The maximum marginal log-likelihood was -1615.0 for Model 1, -1612.6 for Model 2, -1603.4 for Model 3, and -1613.9 for Model 4 (see Table 1). It can be seen that Model 3 has the largest value of the maximum marginal log-likelihood or the smallest deviance (minus twice log-likelihood). This indicates that Model 3 fitted the data better than Model 1, Model 2, and Model 4. Because Model 1, Model 2, Model 3, and Model 4 are not nested within each other, the traditional likelihood ratio test cannot be applied to assess differences in model fit among these four models. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were used to select among competing two-part models. When comparing competing models, we selected the model with the lowest information criterion based on the AIC or BIC as the model demonstrating better fit than its comparison models. The AIC was 3253.9 for Model 1, 3249.2

for Model 2, 3230.8 for Model 3, and 3253.8 for Model 4 (see Table 1). The BIC was 3295.9 for Model 1, 3291.2 for Model 2, 3272.8 for Model 3, and 3299.3 for Model 4. It can be seen that Model 3 had the smallest AIC and BIC. Both the AIC and BIC criteria favor Model 3 over Model 1, Model 2, and Model 4. Based on the results from the information criteria, the maximum marginal likelihood, and Q-Q plots of Pearson residuals, we conclude that Model 3, a two-part logit-gamma random effects model, is the best fit model.

We next fitted data to Model 1, Model 2, Model 3, and Model 4 with the logit link function in Part 1 replaced by the probit link function given in (2) to examine whether prediction accuracy of drinking from a bottle in Part 1, the parameter estimates in Part 2, and the estimated cross-part correlation coefficient vary with the link function used in Part 1. The parameter estimates and their corresponding standard errors for the four models are given in Table 2. The four models gave nearly identical estimates for the parameters and random intercept variance  $\tau^2$  in Part 1. There were no changes in the sign of the parameter estimates for predicting the probability of drinking from a bottle compared to the same models with the logit link function in Part 1 given in Table 1. ROC curves for the probit link function in Part 1 for the four models are given in Figure 4. Figure 4 indicates that the four models have nearly identical ROC curves. Like the ROC curves for the logit link function in Part 1, all the curves followed closely the left-hand border and then the top border of the ROC space. The estimated area under the ROC curve was 0.9796 for Model 1, 0.9796 for Model 2, 0.9796 for Model 3, and 0.9798 for Model 4, with the corresponding standard errors given by 0.00453, 0.00454, 0.00456, and 0.00446. The estimated area under the ROC curve for the probit link in Part 1 is nearly identical to that for the logit link in Part 1 for all four models. Comparisons of the parameter estimates in Part 2 and the estimated correlation coefficient between two random intercepts given in Table 2 with those in Part 2 given in Table 1 indicate that there are no changes in these parameter estimates and their significance levels for all four models. We conclude from these results that prediction accuracy does not vary with the link function used in Part 1, and the parameters estimates in Part 2 and the estimated cross-part correlation coefficient are not sensitive to the link function used in Part 1.

Table 2: Parameter estimates and standard errors for the four two-part random effects models with the logit link function replaced by the probit link function fitted to the total daily milk intake from bottles data

Variable	Model 1 <sup>a</sup>	Model 2 <sup>a</sup>	Model 3 <sup>a</sup>	Model 4 <sup>a</sup>
Part 1: drinking from a bottle				
Intercept	2.41(0.34)***	2.40(0.34)***	2.41(0.34)***	2.43(0.34)***
Baseline daily milk intake from bottles <sup>b</sup>	0.04(0.01)*	0.04(0.01)*	0.04(0.01)*	0.04(0.01)*
Intervention	-0.76(0.29)*	-0.75(0.29)*	-0.75(0.29)*	-0.78(0.30)**
Months post baseline	-0.22(0.03)***	-0.22(0.03)***	-0.22(0.03)***	-0.22(0.03)***
Part 2: modeling nonzero data				
Intercept	2.94(0.07)***	21.60(1.22)***	3.06(0.07)***	11.50(1.53)***
Baseline daily milk intake from bottles <sup>b</sup>	0.01(0.003)**	0.23(0.06)**	0.01(0.003)**	0.16(0.05)**
Intervention	-0.20(0.07)**	-3.84(1.29)**	-0.19(0.06)**	-3.16(0.98)**
Months post baseline	-0.03(0.01)**	-0.69(0.16)***	-0.03(0.01)**	-0.49(0.15)**
$\alpha$	-	-	5.29(0.53)***	-
$\delta$	-	-	-	12.12(1.19)***
$\sigma$	0.45(0.02)***	7.87(0.47)***	-	3.34(0.90)**
$\tau^2$	3.00(0.84)**	3.00(0.84)**	2.99(0.83)**	3.07(0.86)**
$\nu^2$	0.09(0.03)**	33.94(9.17)**	0.07(0.02)**	14.62(5.97)*
$\rho$	0.59(0.16)**	0.61(0.16)**	0.63(0.17)**	0.73(0.21)**
log-likelihood	-1614.9	-1612.6	-1603.4	-1613.8
AIC	3253.7	3249.1	3230.7	3253.6
BIC	3295.7	3291.1	3272.7	3299.1

<sup>a</sup>Model 1: a two-part probit-log-normal random effects model; Model 2: a two-part probit-truncated normal random effects model; Model 3: a two-part probit-gamma random effects model; Model 4: a two-part probit-skew normal random effects model.

<sup>b</sup>Baseline total daily milk intake from bottles was centered at median.

\*p-value < 0.05; \*\*p-value < 0.01; \*\*\*p-value < 0.0001.

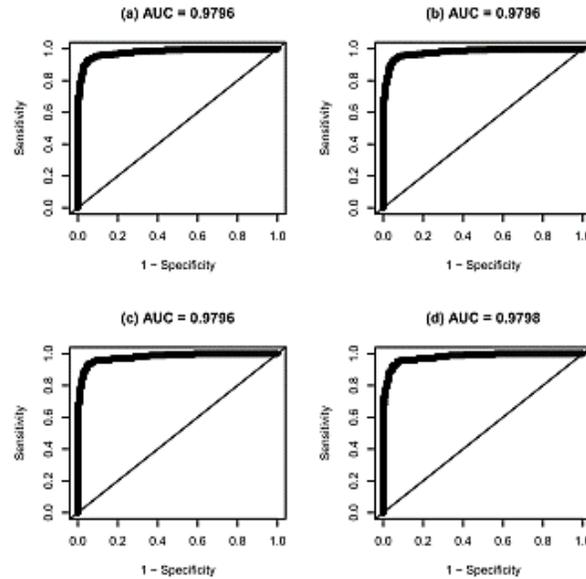


Figure 4: The receiver operating characteristic curves for the probit link function in Part 1 for (a) Model 1: a two-part probit-log-normal random effects model, (b) Model 2: a two-part probit-truncated normal random effects model, (c) Model 3: a two-part probit-gamma random effects model, and (d) Model 4: a two-part probit-skew normal random effects model. AUC stands for the area under the curve.

#### 4. Discussion

We have demonstrated the use of two-part random effects models to assess effects of the bottle-weaning intervention on reducing bottle use and daily milk intake from bottles when the longitudinal response variable has a point mass at 0 and a continuous right-skewed distribution for positive values. We show in all four two-part random effects models (Model 1 to Model 4) that the bottle-weaning intervention promoted bottle-weaning and reduced daily milk intake from bottles in toddlers not off bottles adjusted for baseline total daily milk intake from bottles and age in children aged 11 to 13 months. We also show that there are no differences in model fit using either the logit link function or the probit link function for modeling the probability of drinking from a bottle in Part 1. Prediction accuracy of the logit or probit link function in Part 1 is not sensitive to the distribution assumption on the response variable in Part 2. The parameter estimates in Part 2 and the estimated cross-part correlation coefficient do not vary with the link function used in Part 1. We conclude that Model 3, a two-part logit-gamma random effects model, provides the best overall description of the data compared to the other three two-part models based on the results from the information criteria, the maximum marginal likelihood, and assessing model fit in Part 1 and Part 2 for each of the four competing model. However, it is worth noting that the results of Model 3 being preferred in the current study do not imply that one should favor

Model 3 over the other three models for other repeated measures semi-continuous data sets without thorough assessment of model fit. We have applied Model 1 to Model 4 to fit data on total milk intake from bottles. However, we do not mean to imply that these four models are the only approaches to model repeated measures semi-continuous data. Other distributions such as Weibull or beta (Blough *et al.*, 1999) may be used to model the positive part of a semi-continuous variable. As in some cases, Model 1, Model 2, Model 4, or other models may provide a better fit, the advocacy of one particular model for semi-continuous data either in particular data sets or in general, is not tenable. We suggest fitting all competing two-part models to the data being analyzed and then selecting the model with the best fit, unless there is prior information or expert knowledge on choosing certain two-part models. For example, if positive values are known to be skewed with variance proportional to the mean, gamma regression (Model 3) or log-normal regression (Model 1) can be the choice of the second part model. If positive values are percentages or proportions, beta regression can be the choice of the second part model.

We have fitted the data to two-part models under the assumption that once toddlers are weaned, they will not return to bottle use. This assumption is supported by our data in which all weaned toddlers did not return to bottle use during the study. If some proportion of weaned toddlers return to bottle use during the study, an observation that has no daily milk intake from bottles can be considered as either a true zero from a toddler off bottles or a partial observation censored at 0 from a toddler who drinks from a bottle but happens to not drink from a bottle in the current period. Without further assumptions, two-part models cannot differentiate the true weaned toddlers from those who happen not to use bottles during the study period. Following Moulton and Halsey (1995), the probability of a true zero can be modeled with logistic or probit regression and the other values can be modeled with censored log-normal regression or censored skew normal regression.

## Acknowledgements

The author would like to thank the referee for very helpful comments and suggestions that improved the presentation of this paper.

## References

- [1] Barr, D.R. and Sherrill, E.T. (1999). Mean and variance of truncated normal distributions. *The American Statistician* **53**, 357-361
- [2] Berk, K.N. and Lachenbruch, P.A. (2002). Repeated measures with zeros. *Statistical Methods in Medical Research* **11**, 303-316..
- [3] Blough, D.K., Madden, C.W. and Hornbrook, M.C. (1999). Modeling risk using generalized linear models. *Journal of Health Economics* **18**, 153-171.

- [4] Bonuck, K.A., Huang, V. and Fletcher, J. (2010). Inappropriate bottle use: an early risk for overweight? Literature review and pilot data for a bottle-weaning trial. *Maternal & Child Nutrition* **164**, 1098-1104.
- [5] Bonuck, K.A. and Kahn, R. (2002). Prolonged bottle use and its association with iron deficiency anemia and overweight: a preliminary study. *Clinical Pediatrics* **41**, 603-607.
- [6] Brotanek, J.M., Halterman, J.S., Auinger, P., Flores, G. and Weitzman, M. (2005). Iron deficiency, prolonged bottle-feeding, and racial/ethnic disparities in young children. *Archives of Pediatrics & Adolescent Medicine* **159**, 1038-1042.
- [7] Chai, H.S. and Bailey, K.R. (2008). Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero. *Statistics in Medicine* **27**, 3643-3655.
- [8] Cragg, J.G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* **39**, 829-844.
- [9] Duan, N., Manning, W.G., Morris, C.N. and Newhouse, J.P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics* **1**, 115-126.
- [10] Gooze, R.A., Anderson, S.E. and Whitaker, R.C. (2011). Prolonged bottle use and obesity at 5.5 years of age in US children. *Journal of Pediatrics* **159**, 431-436.
- [11] Hyden, C., Kahn, R. and Bonuck, K. (2012). Bottle-weaning intervention tools: the "how" and "why" of a WIC-based educational flipchart, parent brochure, and web site. *Health Promotion Practice*, April 5.
- [12] Lee, A.H. and Xiang, L. (2011), Mixture analysis of heterogeneous physical activity outcomes. *Annals of Epidemiology* **21**, 780-786.
- [13] Lee, A.H., Zhao, Y., Yau, K.K. and Xiang, L. (2010). How to analyze longitudinal multilevel physical activity data with many zeros? *Preventive Medicine* **51**, 476-481.
- [14] Li, N., Elashoff, D.A., Robbins, W.A. and Xun, L. (2011). A hierarchical zero-inflated log-normal model for skewed responses. *Statistical Methods in Medical Research* **20**, 175-189.
- [15] Liu, L., Cowen, M.E., Strawderman, R.L. and Shih, Y.C.T. (2010). A flexible two-part random effects model for correlated medical costs. *Journal of Health Economics* **29**, 110-123.

- 
- [16] Mahmud, S., Lou, W.Y.W. and Johnston, N.W. (2010). A probit-log-skew-normal mixture model for repeated measures data with excess zeros, with applications to a cohort study of pediatric respiratory symptoms. *BMC Medical Research Methodology* **10**, 55.
- [17] Mills, E.D. (2013). *Adjusting for covariates in zero-inflated gamma and zero-inflated log-normal models for semicontinuous data*. Dissertation. The University of Iowa.
- [18] Moulton, L.H. and Halsey, N.A. (1995). A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics* **51**, 1570-1578.
- [19] Olsen, M.K. and Schafer, J.L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of American Statistical Association* **96**, 730-745.
- [20] Poirier, DJ. (1978). The use of the Box-Cox transformation in limited dependent variable models. *Journal of the American Statistical Association* **73**, 284-287.
- [21] Powell J.L. (1986). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica* **54**, 1435-1460.
- [22] Saei, A., Ward, J. and McGilchrist, C. A. (1996). Threshold models in a methadone programme evaluation. *Statistics in Medicine* **15**, 2253-2260.
- [23] Su, L.I., Brian, D.M.T. and Farewell, V.T. (2009). Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics* **10**, 374-389.
- [24] Sutcliffe, T.L., Khambalia, A., Westergard, S., Jacobson, S., Peer, M. and Parkin, P.C.(2006). Iron depletion is associated with daytime bottle-feeding in the second and third years of life. *Archives of Pediatrics & Adolescent Medicine* **160**, 1114-1120.
- [25] The American Academy of Pediatrics Committee on Nutrition. (1992). The use of whole cow's milk in infancy. *Pediatrics* **89**, 1105-1109.
- [26] Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24-36.
- [27] Tooze, J.A., Grunwald, G.K. and Jones, R.H. (2002). Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research* **11**, 341-355.

Received March 15, 2013; accepted November 10, 2013.

Albert Einstein College of Medicine  
1300 Morris Park Avenue, Bronx, New York 10461, U.S.A.  
yungtai.lo@einstein.yu.edu