

A marginalized two-part model for semicontinuous data

Valerie A. Smith,^{a,b,*†} John S. Preisser,^b Brian Neelon^{a,c}
and Matthew L. Maciejewski^{a,d}

In health services research, it is common to encounter semicontinuous data characterized by a point mass at zero followed by a right-skewed continuous distribution with positive support. Examples include health expenditures, in which the zeros represent a subpopulation of patients who do not use health services, while the continuous distribution describes the level of expenditures among health services users. Semicontinuous data are typically analyzed using two-part mixture models that separately model the probability of health services use and the distribution of positive expenditures among users. However, because the second part conditions on a non-zero response, conventional two-part models do not provide a marginal interpretation of covariate effects on the overall population of health service users and non-users, even though this is often of greatest interest to investigators. Here, we propose a marginalized two-part model that yields more interpretable effect estimates in two-part models by parameterizing the model in terms of the marginal mean. This model maintains many of the important features of conventional two-part models, such as capturing zero-inflation and skewness, but allows investigators to examine covariate effects on the overall marginal mean, a target of primary interest in many applications. Using a simulation study, we examine properties of the maximum likelihood estimates from this model. We illustrate the approach by evaluating the effect of a behavioral weight loss intervention on health-care expenditures in the Veterans Affairs health-care system. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: health-care expenditures; log-skew-normal distribution; marginalized models; semicontinuous data; two-part model; weight loss intervention

1. Introduction

In health services research, it is common to encounter semicontinuous data, such as medical expenditures [1,2], which are characterized by a point mass at zero followed by a right-skewed continuous distribution with positive support. In the case of medical expenditures, the point mass at zero represents a population of ‘non-users’ who do not receive medical care in a given time interval and therefore have no medical expenditures; the continuous distribution, on the other hand, represents the level of expenditures among health services users given that expenditures were incurred. Considering the two defining components of such outcomes, semicontinuous data can be viewed as arising from two distinct stochastic processes: one governing the occurrence of zeros and the second determining the observed value conditional on it being a non-zero response. The first process is commonly referred to as the ‘occurrence’ or ‘binary’ part of the data, while the second is often termed the ‘intensity’ or ‘continuous’ part. Other examples of semicontinuous outcomes include hospital length of stay [3], health assessment scores [4], and average daily alcohol consumption [5,6].

There is extensive literature describing two-part mixture models for analyzing semicontinuous data. Aitchison initially highlighted the need for these two defining processes for unbiased estimation in applications involving estimation of expenditures and number of children per household [7]. Deriving

^aCenter for Health Services Research in Primary Care, Durham VAMC, Durham NC, U.S.A.

^bDepartment of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420, U.S.A.

^cDepartment of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham NC, U.S.A.

^dDivision of General Internal Medicine, Department of Medicine, Duke University Medical Center, Durham NC, U.S.A.

*Correspondence to: Valerie A. Smith, Department of Biostatistics, University of North Carolina, 135 Dauer Drive, 3101 McGavran-Greenberg Hall, CB 7420, Chapel Hill, NC 27599-7420 U.S.A.

†E-mail: vasmith@bios.unc.edu

semicontinuous counterparts to many commonly used probability distributions, he defined these distributions as a mixture of the binary stochastic process and the continuous positive-valued process conditional on observing a positive response. Cragg [8], Manning and Duan [1, 2], and others extended this approach to the regression setting, modeling the binary and continuous components as functions of covariates. Most commonly, the binary part is modeled via logistic regression and the continuous component via a log-normal model. However, because the log-normal distribution imposes a sometimes unrealistic condition of symmetry on the log scale, alternative distributions such as the log-skew-normal (LSN) have recently been proposed for the continuous part in an effort to relax these somewhat restrictive assumptions [9, 10]. More recent extensions include incorporating longitudinal data [5, 11], assessing bias [4], and examining alternative data transformations [12].

Because they explicitly accommodate both data generating processes, two-part mixture models are an ideal choice for modeling semicontinuous data. When adjusting for covariates, these models typically include one set of parameters for the binary response and a second set for the continuous component conditional on a positive response. In particular, covariates in the second, or continuous, part are interpreted conditionally upon having observed a positive outcome. Consequently, attempts to combine these two parts to form the overall marginal mean effect of any covariate relies on specifying values for each of the other covariates in the model. As such, it is generally challenging to obtain a straightforward interpretation of covariate effects on the marginal mean in two-part models.

In many cases, however, investigators' main interest lies in examining such effects on the marginal mean in order to draw conclusions about the impact of predictors on the population as a whole. For example, in economic studies of system-wide health-care expenditures, investigators and policy makers may wish to understand the average effect on medical expenditures of increasing specialty care copayments [13] or of bariatric surgery for weight loss [14, 15] on the entire affected or eligible populations rather than estimating separate effects for the probability of incurring expenditures and the level of expenditures given that any are incurred. In particular, an intervention may have one effect on the probability of occurrence but the opposite effect on the intensity given occurrence. In such cases, policy makers may be left without a true understanding of the overall population-level effect of such an intervention.

To achieve more interpretable effects, Mullahy [12] and Buntin and Zaslavsky [16] proposed using a one-part exponential conditional mean model to estimate effects of covariates on the marginal mean. While this one-part model provides interpretable estimates, it does not explicitly account for the zero-inflated nature of the data or provide investigators with estimates of covariate effects on the probability of occurrence. Thus, alternative models must be considered when interest lies in estimating both the binary component and the overall marginal mean.

We propose a new 'marginalized' two-part model for semicontinuous data, which yields more interpretable effect estimates in two-part models by parameterizing the model in terms of the marginal mean. This model maintains many of the important features of conventional two-part models, such as capturing zero-inflation and skewness, but allows investigators to examine covariate effects on the overall marginal mean, a target of primary interest in many applications. We also propose an extension to accommodate LSN data to relax the commonly used log-normal assumption for the continuous part of the model. We illustrate the approach by evaluating the effect of a behavioral weight loss intervention on health-care expenditures in the Veterans Affairs (VA) health-care system.

The remainder of the paper is organized as follows: Section 2 introduces the marginalized two-part (MTP) log-normal model and extends it to an LSN distribution. Section 3 presents results from a simulation study highlighting important features of our method. Section 4 applies the approach to the behavioral weight loss program, and Section 5 provides a discussion and points to areas for future work.

2. Marginalized two-part models for semicontinuous data

2.1. Conventional two-part model

We begin with a review of the conventional two-part model presented in Cragg [8], Manning and Duan [1, 2], and elsewhere. For data consisting of independent observations, the generic form of the conventional two-part model can be written as

$$f(y_i) = (1 - \pi_i)^{1_{(y_i=0)}} \times [\pi_i g(y_i | y_i > 0)]^{1_{(y_i > 0)}}, \quad y_i \geq 0, \quad i = 1, \dots, n \quad (1)$$

where $\pi_i = \Pr(Y_i > 0)$, $1_{(\cdot)}$ is the indicator function, and $g(y_i|y_i > 0)$ is any density function applicable to the positive values of Y_i , although the log-normal density is often chosen. This model is parameterized as

$$\text{logit}(\pi_i) = \mathbf{z}_i' \boldsymbol{\alpha} \quad \text{and} \quad (2)$$

$$\mu_i = E(\ln Y_i | Y_i > 0) = \mathbf{x}_i' \boldsymbol{\gamma} \quad (3)$$

When fitting this model to independent responses, the binary and conditionally continuous components of the likelihood are separable, and therefore, these two parts are fit separately. The binary component is often modeled using logistic regression, and the continuous component can be fit using standard regression models, such as the log-normal.

2.2. Marginalized two-part model

To obtain interpretable covariate effects on the marginal mean, we propose the following *MTP model* that parameterizes the covariate effects directly in terms of the marginal mean, $v_i = E(Y_i)$, on the original (i.e., untransformed) data scale. The MTP model specifies the linear predictors

$$\text{logit}(\pi_i) = \mathbf{z}_i' \boldsymbol{\alpha} \quad \text{and} \quad (4)$$

$$E(Y_i) = v_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}) \quad (5)$$

Parameter estimates can be obtained using standard optimization routines such as Newton–Raphson or Fisher scoring. Model-predicted means and standard errors can also be easily obtained under this parameterization in a single step by estimating $\exp(\mathbf{x}_i' \boldsymbol{\beta})$ at the desired values of the covariates.

2.3. Comparison of treatment effect estimates

Using the conventional model shown in Equation (3), γ_j is interpreted as the effect of a unit increase in the j th covariate, x_{ij} , on the conditional mean of $\ln(Y_i)$ given Y_i is positive. In many applications, however, this interpretation has limited usefulness as it is only relevant for the population of health services users. Rather, interest often lies in estimating the effect of covariates \mathbf{x}_i on the marginal mean of Y_i for the combined population of health services users and non-users; that is the effect of \mathbf{x}_i on $E(Y_i)$ unconditionally. In the case of the log-normal distribution, that is the effect of \mathbf{x}_i on

$$E(Y_i) = v_i = \pi_i \exp(\mu_i + \sigma^2/2) = \frac{e^{\mathbf{x}_i' \boldsymbol{\alpha}}}{1 + e^{\mathbf{x}_i' \boldsymbol{\alpha}}} \exp(\mathbf{x}_i' \boldsymbol{\gamma} + \sigma^2/2) \quad (6)$$

where σ^2 is the variance of Y_i on the log scale. Assuming $\mathbf{x}_i = \mathbf{z}_i$, as is commonly specified, it follows from (6) that the per-unit effect of the j th covariate, x_{ij} , on the marginal mean is

$$\frac{E(Y_i | x_{ij} = j + 1, \tilde{\mathbf{x}}_i)}{E(Y_i | x_{ij} = j, \tilde{\mathbf{x}}_i)} = \frac{1 + \exp[\tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\alpha}} + \alpha_j \cdot j]}{1 + \exp[\tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\alpha}} + \alpha_j \cdot (j + 1)]} \exp(\alpha_j + \gamma_j) \quad (7)$$

where $\tilde{\mathbf{x}}_i$ is \mathbf{x}_i with x_{ij} removed and $\tilde{\boldsymbol{\alpha}}$ is $\boldsymbol{\alpha}$ with α_j removed. Thus, unless $\alpha_j = 0$, one must specify fixed values for the remaining covariates in order to obtain a marginal interpretation for the effect of x_{ij} . Further, to obtain confidence intervals (CIs) or formal inference on this marginal effect, the delta method or resampling techniques must be employed.

Using the MTP model as parameterized in Equation (5), $\boldsymbol{\beta}$ is estimated for the entire population as opposed to $\boldsymbol{\gamma}$, which is conditional on $Y_i > 0$. Unlike γ_j in the conventional model, $\exp(\beta_j)$ can be interpreted as the multiplicative effect on the unconditional marginal mean, v_i , when covariate x_{ij} increases by one unit. In other words, the left-hand side of (7) equals $\exp(\beta_j)$ under model (5). Unlike the conventional model, standard errors and CIs for covariate effects on the marginal mean are easily obtained as part of the standard model output.

There are other important distinctions between the models. In particular, when the model includes ancillary covariates with no interactions, the MTP model assumes a homogeneous treatment effect on

$E(Y_i)$, whereas the conventional model yields heterogeneous effects that depend on the specific values of the additional covariates, potentially creating misleading results. As an illustrative example, we generated a simulated dataset of sample size 10,000 using the following specification:

$$\begin{aligned}\text{logit}(\pi_i) &= \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} \quad \text{and} \\ E(Y_i) &= v_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})\end{aligned}$$

where $x_{i1} \sim N(50, 100)$ and $x_{i2} \sim \text{Bernoulli}(0.5)$. We specified parameters values as $\alpha_0 = 14.4$, $\alpha_1 = -0.3$, $\alpha_2 = 1.6$, $\beta_0 = 5$, $\beta_1 = 0.05$, and $\beta_2 = 1.1$ and assumed that Y_i followed the standard two-part distribution given in Equation (1) with a log-normal density specified for $g(y_i|y_i > 0)$. Under this scenario, the true multiplicative ‘treatment effect’ on $E(Y_i)$ of x_{i2} taking a value of 1 versus 0 is $\exp(\beta_2) = \exp(1.1) = 3.0$. Because the ratio $E(Y_i|x_{i2} = 1, x_{i1})/E(Y_i|x_{i2} = 0, x_{i1}) = \exp(\beta_2)$ does not rely on specification of other covariate values, note that this treatment effect is the same regardless of the values for x_{i1} . Next, using Equation (7), we estimated treatment effects under the conventional model at the first, second, and third quartiles of x_{i1} , taking values 43, 50, and 57, respectively. Under the conventional model, the estimated effects of an increase in x_{i2} on $E(Y_i)$ were multiplicative increases of 2.5, 4.4, and 8.3, respectively, while the true multiplicative increase was 3.0 regardless of the value of x_{i1} . It is worth pointing out that when treatment effect heterogeneity truly exists, the MTP model can accommodate this through the systematic inclusion of interactions, which should be driven by subject-matter considerations. While the conventional model also accommodates the systematic inclusion of subject-matter driven interactions, it imposes an arbitrary heterogeneity that always exists unless one omits the treatment covariate from the binary part of the model.

Because the interpretation of the marginal treatment effect from the conventional model relies on specified values of each other covariate in the model and includes arbitrary heterogeneity, ‘standardization’ [17] is often used to obtain marginal treatment effect estimates by averaging across the observed heterogeneous estimates. In this approach, after modeling is complete, expected outcomes are estimated for each individual in the sample by first assuming they were a member of the treatment group then assuming they were a member of the control group; the overall treatment effect is then estimated as the difference in means of these constructed treatment and control groups. This approach has several disadvantages, however. First, bootstrapping or other resampling techniques must be employed to obtain standard errors and CIs for estimated treatment effects. Further, the effect estimates are averaged over the sample distribution of the observed values of the other covariates in the model. Thus, if the sample distribution of these covariates does not represent the distribution in the target population, the treatment effect estimate is sample specific and not as easily generalized to the overall population. This is not likely to be a problem in large datasets where the sample covariate distributions accurately represent those in the overall population. However, in smaller samples, the covariate distributions may not be representative of the population as a whole, leading to biased inferences. Additionally, estimates obtained via standardization lack generalizability to other populations that vary in the distribution of these covariates. As a result, the conventional two-part model lacks appeal when the objective is to estimate the effect of a covariate on the marginal mean. When using the MTP model, however, estimation of the treatment effect does not require averaging over the observed values of the other covariates in the sample; therefore, these estimates provide much greater generalizability and ease of computation.

This is not to say that the MTP model should be preferred over the conventional model in all cases. Indeed, when the primary target of inference is $E(Y_i|Y_i > 0)$, the MTP model engenders arbitrary heterogeneity and provides less interpretable estimates on the conditional mean of Y among the positive values. Ultimately, the choice between models should be guided by the aims of the analyst. If the aim is to model treatment effects on $E(Y_i)$ in the presence of confounders, one should use the MTP model; on the other hand, if the target of inference is $E(Y_i|Y_i > 0)$, the conventional model should be used.

2.4. Marginalized two-part log-normal model

When modeling semicontinuous data, the continuous component is most frequently modeled using a log-normal distribution. The generic form of the two-part log-normal model for independent responses can be written as in (1) with $g(y_i|y_i > 0)$ taking the log-normal density function $\text{LN}(\cdot; \mu, \sigma^2)$ with mean μ and variance σ^2 on the log scale. The marginal mean and variance of Y_i are then given by [7]:

$$E(Y_i) = v_i = \pi_i \exp(\mu_i + \sigma^2/2) \quad (8)$$

$$\text{Var}(Y_i) = \pi_i \exp(2\mu_i + \sigma^2) [\exp(\sigma^2) - \pi_i] \quad (9)$$

The likelihood, parameterized in terms of π_i and μ_i , is

$$\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\mu} | \mathbf{y}) = \prod_i (1 - \pi_i)^{1_{(y_i=0)}} \left\{ \frac{\pi_i}{y_i \sqrt{2\pi\sigma}} \exp \left[-\frac{1}{2\sigma^2} (\ln y_i - \mu_i)^2 \right] \right\}^{1_{(y_i>0)}}$$

In order to utilize this log-normal likelihood framework, the marginal mean in Equation (8) can be rearranged to solve for μ_i , yielding

$$\begin{aligned} \mu_i &= \ln y_i - \ln \pi_i - \sigma^2/2 \\ &= \mathbf{x}_i' \boldsymbol{\beta} - \ln \pi_i - \sigma^2/2 \end{aligned}$$

Noting that

$$\begin{aligned} \pi_i &= \frac{e^{z_i' \boldsymbol{\alpha}}}{1 + e^{z_i' \boldsymbol{\alpha}}} \Rightarrow \ln \pi_i = z_i' \boldsymbol{\alpha} - \ln(1 + e^{z_i' \boldsymbol{\alpha}}), \text{ and} \\ \ln(1 - \pi_i) &= -\ln(1 + e^{z_i' \boldsymbol{\alpha}}) \end{aligned}$$

we can express the log-likelihood in terms of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ :

$$\begin{aligned} l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma) &= \sum_i -\ln(1 + e^{z_i' \boldsymbol{\alpha}}) + \sum_{y_i>0} \left\{ z_i' \boldsymbol{\alpha} - \ln y_i - \frac{1}{2} \ln 2\pi - \ln \sigma \right. \\ &\quad \left. - \frac{1}{2\sigma^2} [\ln y_i + z_i' \boldsymbol{\alpha} - \ln(1 + e^{z_i' \boldsymbol{\alpha}}) + \sigma^2/2 - \mathbf{x}_i' \boldsymbol{\beta}]^2 \right\} \end{aligned}$$

with score equations

$$\mathbf{U}_i = \left[\frac{\partial l_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)}{\partial \boldsymbol{\alpha}} \quad \frac{\partial l_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)}{\partial \boldsymbol{\beta}} \quad \frac{\partial l_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)}{\partial \sigma} \right]'$$

where

$$\frac{\partial l_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)}{\partial \boldsymbol{\alpha}} = \left\{ \frac{-e^{z_i' \boldsymbol{\alpha}}}{1 + e^{z_i' \boldsymbol{\alpha}}} + \left[1 - \frac{1}{\sigma^2} [\ln y_i + z_i' \boldsymbol{\alpha} - \ln(1 + e^{z_i' \boldsymbol{\alpha}}) + \frac{1}{\sigma^2} - \mathbf{x}_i' \boldsymbol{\beta}] \cdot \left(\frac{1}{1 + e^{z_i' \boldsymbol{\alpha}}} \right) \right] 1_{(y_i>0)} \right\} z_i'$$

$$\frac{\partial l_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)}{\partial \boldsymbol{\beta}} = \left\{ \frac{1}{\sigma^2} [\ln y_i + z_i' \boldsymbol{\alpha} - \ln(1 + e^{z_i' \boldsymbol{\alpha}}) - \mathbf{x}_i' \boldsymbol{\beta}] + \frac{1}{2} \right\} \mathbf{x}_i'$$

Note the typo: $s^2/2$ not $1/s^2$

$$\begin{aligned} \frac{\partial l_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)}{\partial \sigma} &= \frac{-1}{\sigma} \left\{ 1 - \frac{1}{\sigma^2} [\ln y_i + z_i' \boldsymbol{\alpha} - \ln(1 + e^{z_i' \boldsymbol{\alpha}}) + \sigma^2/2 - \mathbf{x}_i' \boldsymbol{\beta}]^2 \right. \\ &\quad \left. + \ln y_i + z_i' \boldsymbol{\alpha} - \ln(1 + e^{z_i' \boldsymbol{\alpha}}) + \frac{\sigma^2}{2} - \mathbf{x}_i' \boldsymbol{\beta} \right\} \end{aligned}$$

With the conventional model, the likelihood and score equations can be separated into two independent components: one for the binary part and one for the continuous part. In contrast, note that the score equations for the MTP model are not separable, and thus, the binary and continuous parts are fit simultaneously. Model-based asymptotic standard errors are computed using Fisher's information matrix, $\mathbf{I}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)$ as

$$\text{s.e.}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}) = \sqrt{\text{diag} [\mathbf{I}^{-1}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)]}$$

with the maximum likelihood estimates substituted for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ .

2.5. Extension to the log-skew-normal distribution

While the log-normal distribution is suitable for many outcomes, it requires the somewhat restrictive assumption that the log-transformed outcome is symmetric and normally distributed, an assumption that is often violated in practice. We can relax this assumption by instead selecting for the positive responses an LSN distribution, which accommodates skewness through the inclusion of a shape parameter. Using the same linear predictors as in Equations (4) and (5), the generic form of the two-part LSN model for independent data is given by

$$f(y_i) = (1 - \pi_i)^{1_{(y_i=0)}} \times [\pi_i \text{LSN}(y_i; \xi_i, \omega, \kappa)]^{1_{(y_i>0)}}, \quad y_i \geq 0, \quad i = 1, \dots, n$$

where $\text{LSN}(\cdot; \xi_i, \omega, \kappa)$ denotes the LSN distribution with location parameter ξ_i , scale parameter $\omega > 0$, and shape parameter κ , all on the log scale, given by

$$g(y_i | y_i > 0) = \frac{2}{\omega y_i} \phi\left(\frac{\ln y_i - \xi_i}{\omega}\right) \Phi\left(\frac{\kappa}{\omega}(\ln y_i - \xi_i)\right)$$

The marginal mean and variance of Y_i for the MTP LSN model are given by

$$\begin{aligned} E(Y_i) &= v_i = 2\pi_i \exp\left(\xi_i + \frac{\omega^2}{2}\right) \Phi(\omega\delta) \\ \text{Var}(Y_i) &= 2\pi_i \exp(2\xi_i + \omega^2) [\exp(\omega^2) \Phi(2\omega\delta) - 2\pi_i (\Phi(\omega\delta))^2] \end{aligned}$$

where $\delta = \frac{\kappa}{\sqrt{1+\kappa^2}}$ and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal density. The likelihood, parameterized in terms of $\boldsymbol{\pi}$ and $\boldsymbol{\xi}$, is then

$$\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\xi}, \omega, \kappa) = \prod_i (1 - \pi_i)^{1_{(y_i=0)}} \left\{ \frac{2\pi_i}{\omega y_i} \phi\left(\frac{\ln y_i - \xi_i}{\omega}\right) \Phi\left(\frac{\kappa}{\omega}(\ln y_i - \xi_i)\right) \right\}^{1_{(y_i>0)}}$$

where $\phi(\cdot)$ is the probability density function for the standard normal distribution. Thus, the log-likelihood in terms of $\boldsymbol{\pi}$ and $\boldsymbol{\xi}$ is

$$\begin{aligned} l(\boldsymbol{\pi}, \boldsymbol{\xi}, \omega, \kappa) &= \sum_{y_i=0} \ln(1 - \pi_i) + \sum_{y_i>0} \left\{ \ln \pi_i + \ln 2 - \ln(\omega) - \ln(y_i) + \ln \left[\phi\left(\frac{\ln y_i - \xi_i}{\omega}\right) \right] \right. \\ &\quad \left. + \ln \left[\Phi\left(\frac{\kappa}{\omega}(\ln y_i - \xi_i)\right) \right] \right\} \end{aligned} \quad (10)$$

In order to re-express the LSN likelihood as a function of $\boldsymbol{\beta}$, we first solve for ξ_i in terms of $\boldsymbol{\beta}$:

$$\begin{aligned} \xi_i &= \ln v_i - \ln 2 - \ln \pi_i - \ln [\Phi(\omega\delta)] - \frac{\omega^2}{2} \\ &= \mathbf{x}'_i \boldsymbol{\beta} - \ln 2 - \ln \pi_i - \ln [\Phi(\omega\delta)] - \frac{\omega^2}{2} \end{aligned}$$

Plugging this into Equation (10), the log-likelihood expressed in terms of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, ω , and κ is

$$\begin{aligned} l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \omega, \kappa) &= \sum_i -\ln(1 + e^{z_i' \boldsymbol{\alpha}}) + \sum_{y_i>0} \left\{ \mathbf{z}'_i \boldsymbol{\alpha} + \ln 2 - \ln \omega - \ln(y_i) \right. \\ &\quad \left. + \ln \left[\phi\left(\frac{1}{\omega} \left(\ln y_i - \mathbf{x}'_i \boldsymbol{\beta} + \ln 2 + \mathbf{z}'_i \boldsymbol{\alpha} - \ln(1 + e^{z_i' \boldsymbol{\alpha}}) + \ln(\Phi(\omega\delta)) + \frac{\omega^2}{2} \right) \right) \right] \right. \\ &\quad \left. + \ln \left[\Phi\left(\frac{\kappa}{\omega} \left(\ln y_i - \mathbf{x}'_i \boldsymbol{\beta} + \ln 2 + \mathbf{z}'_i \boldsymbol{\alpha} - \ln(1 + e^{z_i' \boldsymbol{\alpha}}) + \ln(\Phi(\omega\delta)) + \frac{\omega^2}{2} \right) \right) \right] \right\} \end{aligned}$$

Because the LSN reduces to the log-normal model when $\kappa = 0$, the choice between the log-normal model and the LSN model can be easily assessed using a likelihood ratio test. It should be noted that the LSN likelihood can be somewhat more sensitive to starting values with small sample sizes (say, $n < 30$) when using software such as SAS PROC NLMIXED (SAS Institute, Cary, NC, USA) that require initial values to be prespecified. This likelihood also has a stationary point at $\kappa = 0$ [9]; as such, 0 should not be provided as a starting value for κ in such estimation routines.

3. Simulation study

To assess the performance of our proposed MTP model, we generated simulated data using the following specification:

$$\text{logit}(\pi_i) = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} \quad \text{and} \\ E(Y_i) = v_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})$$

where $x_{i1} \sim N(50, 100)$ and $x_{i2} \sim \text{Bernoulli}(0.5)$. We specified parameters values as $\alpha_0 = 14.4$, $\alpha_1 = -0.3$, $\alpha_2 = 1.6$, $\beta_0 = 5$, $\beta_1 = 0.05$, and $\beta_2 = 1.1$. Using this specification, we generated 1000 samples of size 10,000 under three scenarios with varying levels of skewness in the distribution:

- (i) $f(y_i) = (1 - \pi_i)^{1_{(y_i=0)}} \times [\pi_i \text{LN}(y_i; \mu_i, \sigma^2)]^{1_{(y_i>0)}}$ with $\sigma^2 = 4$, or equivalently $f(y_i) = (1 - \pi_i)^{1_{(y_i=0)}} \times [\pi_i \text{LSN}(y_i; \xi_i, \omega, \kappa)]^{1_{(y_i>0)}}$ with $\omega = 2$ and $\kappa = 0$;
- (ii) $f(y_i) = (1 - \pi_i)^{1_{(y_i=0)}} \times [\pi_i \text{LSN}(y_i; \xi_i, \omega, \kappa)]^{1_{(y_i>0)}}$ with $\omega = 2$ and $\kappa = 2$; and
- (iii) $f(y_i) = (1 - \pi_i)^{1_{(y_i=0)}} \times [\pi_i \text{LSN}(y_i; \xi_i, \omega, \kappa)]^{1_{(y_i>0)}}$ with $\omega = 2$ and $\kappa = 10$.

Under scenario (i), data were initially generated from a log-normal distribution with mean μ_i as shown in Section 2.4 and variance σ^2 on the log scale. Similarly, under scenarios (ii) and (iii), the data were generated from an LSN distribution with location parameter ξ_i as defined in Section 2.5, scale parameter ω , and shape parameter κ , all on the log scale. Excess zeros were introduced in the Y_i 's with probability π_i , resulting in 48% zeros. The log-normal data were generated using SAS 9.3, and the LSN data were generated using R version 2.15.2 [18] using the SN package [19]. All models were fit using SAS 9.3 NLMIXED. The SAS code for fitting the log-normal and LSN MTP models is provided in the Appendix.

Table I shows the percent relative bias, median standard error, and coverage probability of each parameter from fitting each sample to both the log-normal and LSN MTP models. The log-normal marginalized

Table I. Marginalized two-part model performance with 1000 simulations and varying skewness.

κ	Parameter	True value	Log-normal model			Log-skew-normal model		
			Percent Relative median bias	Median std error	Coverage probability	Percent Relative median bias	Median std error	Coverage probability
0	α_0	14.4	0.09	0.2846	0.9488	0.08	0.2846	0.9486
	α_1	-0.3	-0.06	0.0058	0.9509	-0.05	0.0058	0.9497
	α_2	1.6	-0.08	0.0624	0.9468	-0.07	0.0624	0.9455
	β_0	5.0	0.11	0.1728	0.9478	0.44	0.1774	0.9486
	β_1	0.05	-0.41	0.0039	0.9549	-0.37	0.0039	0.9549
	β_2	1.1	-0.10	0.0585	0.9428	-0.12	0.0585	0.9444
2	α_0	14.4	0.12	0.2738	0.9489	0.13	0.2716	0.9429
	α_1	-0.3	-0.11	0.0055	0.9469	-0.14	0.0055	0.9419
	α_2	1.6	0.23	0.0604	0.9550	0.22	0.0599	0.9539
	β_0	5.0	-4.82	0.1218	0.4865	0.07	0.1222	0.9469
	β_1	0.05	-0.35	0.0028	0.9560	-0.19	0.0027	0.9539
	β_2	1.1	0.22	0.0419	0.9520	0.18	0.0400	0.9419
10	α_0	14.4	0.11	0.2672	0.9459	0.12	0.2298	0.9449
	α_1	-0.3	-0.12	0.0054	0.9429	-0.14	0.0046	0.9469
	α_2	1.6	0.19	0.0591	0.9520	0.05	0.0513	0.9570
	β_0	5.0	-6.81	0.1067	0.1111	0.0007	0.0764	0.9469
	β_1	0.05	-0.26	0.0025	0.9449	-0.06	0.0016	0.9489
	β_2	1.1	-0.08	0.0369	0.9580	0.09	0.0228	0.9309

model failed to converge three times when $\kappa = 0$ and once each when $\kappa = 2$ and $\kappa = 10$; the LSN model failed to converge 46 times when $\kappa = 0$, twice when $\kappa = 2$, and once when $\kappa = 10$. Likelihood ratio tests favored the LSN model in 3.2% of samples when $\kappa = 0$ and in 100% of samples when $\kappa = 2$ or $\kappa = 10$.

Under all scenarios, bias remained small, and coverage probabilities were approximately 0.95 for all parameters except β_0 . As skewness increased, bias increased for β_0 under the log-normal MTP model, and the coverage probability dropped to as low as 0.11 when $\kappa = 10$. While estimates of the remaining parameters would still be valid regardless of which model were used, when the data are skewed, the log-normal model is not appropriate for making predictions or estimating the overall mean, $\exp(\mathbf{x}'_i\boldsymbol{\beta})$, because of the bias in β_0 . Efficiency gains were also observed for the LSN MTP model when skewness was present; standard errors were somewhat smaller under the LSN model than the log-normal model when $\kappa = 2$ and even more so when $\kappa = 10$. These results indicate that the proposed MTP models provide unbiased estimates of regression coefficients. In the presence of skewness, however, the LSN model should be used to improve efficiency and yield unbiased predictions of the marginal mean, $\exp(\mathbf{x}'_i\boldsymbol{\beta})$, because this is a function of β_0 .

4. Analysis of MOVE! intervention data

The VA health-care system implemented a system-wide weight loss intervention (MOVE!) beginning in 2006 to address the high prevalence of obesity among VA patients [20]. The high cost of obesity is well documented [21, 22], and the MOVE! intervention is the first behavioral weight loss program implemented across an entire health system. MOVE! was implemented as an unfunded mandate, so understanding its effect on average health-care expenditures is important to guide program planning and refinement in VA. Assessment of the effect of MOVE! on the marginal mean of the entire VA population is also important for other health-care systems that are also considering the adoption of behavioral interventions for reigning in the increasing costs of obesity. We use our MTP model to provide an estimate of the effect of the MOVE! intervention on the marginal mean expenditures among obese veterans.

The data for this analysis were drawn from a retrospective cohort study of obese VA patients eligible for MOVE! in fiscal years 2006–2009 who were identified from a longitudinal study of the VA cost of obesity. Data were obtained from the VA Corporate Data Warehouse and the VA Outpatient Care File. As a part of a larger study, data were first obtained on all veterans who had received VA services and had a weight measurement in 2002 ($N=3,365,004$). This sample was then stratified into veterans who ever had one or more MOVE! clinic visits in 2006–2009 (MOVE! enrollees) or veterans who did not have a MOVE! clinic visit in this timeframe (non-enrollees).

Veterans were excluded from both cohorts if they were older than 70 years in 2010, had a body mass index (BMI) of less than 30 kg/m² within 30 days of the index date, did not have sex data available, or had contraindications to MOVE! use during the year of MOVE! initiation. Weight loss contraindications that warranted exclusion were central nervous system infections, organic brain syndromes or dementias, anorexia, anterior horn diseases, Huntington's disease, cirrhosis, dialysis, emphysema, neurological disorders, hepatitis, recent transplant surgery, or recent cancer treatment. Patients residing in nursing homes, hospice, or residential or adult day health care were also excluded.

To reduce the non-equivalence of MOVE! enrollees and non-enrollees due to imbalance in observed covariates, MOVE! enrollees and non-enrollees were first matched exactly on sex, race (white or non-white), marital status (married or unmarried), copay status (exempt versus non-exempt), and veterans integrated service network of residence. Then, potential matches were retained on the basis of BMI and comorbidity burden, assessed via the 2002 diagnostic cost group (DCG) score. Only matches with the same integer BMI measure occurring within 7 days of baseline measure of their respective MOVE! enrollee and the same DCG score (closest integer) were retained. The final cohort included 18,214 MOVE! enrollees and 18,214 non-enrollees.

The expenditure outcome of interest was total VA expenditures in the fiscal year following MOVE! initiation and was obtained from the VA Health Economics Resource Center. Expenditures for non-VA services were excluded as this analysis took a VA payer perspective. Total expenditures were inflation-adjusted to 2011 dollars using the general Consumer Price Index (CPI) because medical CPI does not adequately account for technological improvement, quality change, and improved health outcomes [23]. The explanatory variable of primary interest was MOVE! initiation, which could occur any time between October 2005 and September 2009.

Descriptive statistics for the covariates and outcome are shown in Table II. Of note, 17% percent of MOVE! enrollees and 14% percent of non-enrollees had zero health-care expenditures in the year

Table II. Means (SD) for MOVE! data.

	MOVE! enrollees (<i>n</i> = 18,214)	Non- enrollees (<i>n</i> = 18,214)
Covariates		
Age	61 (9.3)	61 (9.3)
BMI	35.2 (3.9)	35.1 (3.9)
DCG Score	0.24 (0.17)	0.24 (0.17)
Outcomes		
% Positive cost	83.2	86.2
Total costs	7005 (18,866)	6542 (18,641)
Total costs among users	8424 (20,398)	7588 (19,878)

BMI, body mass index; DCG, diagnostic cost group.

Table III. Marginalized two-part model results: MOVE! example.

		Log-normal model		Log-skew-normal model	
	Parameter	Parameter estimate	Standard error	Parameter estimate	Standard error
$\Pr(Y_i > 0)$					
Intercept	α_0	2.5653	0.1659	2.5680	0.1658
MOVE! enrollment	α_1	-0.2381	0.0292	-0.2382	0.0292
BMI	α_2	-0.0321	0.0036	-0.0321	0.0036
Age	α_3	0.0080	0.0016	0.0080	0.0016
DCG score	α_4	-0.3454	0.0849	-0.3457	0.0848
$E(Y_i)$					
Intercept	β_0	9.1179	0.0875	9.1142	0.0872
MOVE! enrollment	β_1	0.1749	0.0145	0.1790	0.0145
BMI	β_2	0.0079	0.0019	0.0084	0.0019
Age	β_3	-0.0176	0.0008	-0.0174	0.0008
DCG score	β_4	1.2933	0.0446	1.2946	0.0444
	σ^2	1.4680	0.0118		
	ω			1.4123	0.0183
	κ			0.8426	0.0484

BMI, body mass index; DCG, diagnostic cost group.

Table IV. LSN model-estimated means (standard errors) at quartiles of age, BMI, and DCG Score.

	MOVE! enrollees	Non-enrollees
Age 55 years, BMI 32, DCG 0.11	\$6303 (101)	\$5270 (83)
Age 61 years, BMI 34, DCG 0.22	\$6660 (90)	\$5568 (73)
Age 66 years, BMI 37, DCG 0.32	\$7127 (107)	\$5958 (97)

LSN, log-skew-normal; BMI, body mass index; DCG, diagnostic cost group.

following initiation, yielding standard one-part log-normal or LSN models inappropriate. To assess the effect of MOVE! enrollment on health-care expenditures in the following year, we fit the MTP model:

$$\begin{aligned}\logit(\pi_i) &= \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4} \\ E(Y_i) &= \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4})\end{aligned}$$

where $x_{i1} = 1$ if individual i was enrolled in MOVE! and 0 otherwise, and we additionally adjusted for x_{i2} , x_{i3} , and x_{i4} , individual i 's BMI, age, and DCG score, respectively. We fit this model twice, first assuming a log-normal distribution for the positive-valued observations, then assuming the more flexible LSN distribution. Table III presents the parameter estimates and standard errors from the log-normal and LSN MTP models. Table IV presents model-estimated mean expenditures from the LSN model at the quartile values of age, BMI, and DCG score for MOVE! enrollees and non-enrollees.

The estimates from the two models are quite similar, although a likelihood ratio test indicated that the LSN model was the more appropriate fit ($p < 0.0001$). Both models estimate an odds ratio of $\exp(-0.24) = 0.79$, indicating that the odds of incurring health-care expenditures in the fiscal year following MOVE! enrollment were approximately 21% lower for those enrolled in MOVE! compared with non-enrollees with 95% CI (17%, 26%). Despite the lower probability of incurring expenditures, however, we estimated from the log-normal MTP model that enrollment in MOVE! was associated with $\exp(0.1749) = 1.19$ times higher total health-care expenditures on average in the following fiscal year with 95% Wald-type CI (1.16, 1.23). Similarly, the LSN MTP model estimated that MOVE! enrollment was associated with $\exp(0.1790) = 1.20$ times higher total health-care expenditures on average in the following fiscal year with 95% CI (1.16, 1.23). While expenditures for non-enrollees remained lower than those of MOVE! enrollees, expenditures for both groups trended upward with increasing BMI and DCG score. Note that the estimated means for MOVE! enrollees at each quartile were 1.20 times higher than those of non-enrollees, reflecting the homogeneous model-estimated treatment effect across the distribution of age, BMI, and DCG score.

For comparison, we additionally fit the conventional two-part mixture model to these data. Using the same covariates as in the original analyses, we fit a logistic regression model to estimate the probability of incurring positive expenditures among the 36,428 individuals in our cohort and an LSN model on the subset of 30,847 individuals who had positive expenditures to estimate the level of expenditures conditional on occurrence. Thus, we fit the model:

$$\begin{aligned}\text{logit}(\pi_i) &= \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4} \\ E(\ln Y_i | Y_i > 0) &= \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i3} + \gamma_4 x_{i4}\end{aligned}$$

Table V presents the regression estimates and standard errors from this model. The parameter estimates were similar to those estimated from the MTP model in Table III. This reflects the fact that the percentage of zeros was not large, and hence, the marginal mean was primarily driven by the positive expenditure values. Similar to our proposed model, the logistic regression suggested that MOVE! enrollment was associated with 21% lower odds of incurring health-care costs in the following fiscal year compared with non-enrollees (95% CI [16%, 26%]). In contrast, the LSN model estimated that, conditional on incurring expenditures, those enrolled in MOVE! had 0.22 higher expenditures on the log scale on average than those not enrolled in MOVE! ($p < 0.0001$). However, with such conflicting results in the binary and continuous parts of this model, investigators are left without a clear sense of the combined overall effect of such an intervention on the average population cost. The MTP model, on the other hand, provides a single, easily interpreted estimate of the overall effect.

Table V. Conventional two-part LSN mixture model results: MOVE! example.

	Parameter	Parameter estimate	Standard error
$\Pr(Y_i > 0)$			
Intercept	α_0	2.4906	0.1686
MOVE! enrollment	α_1	-0.2369	0.0293
BMI	α_2	-0.0320	0.0036
Age	α_3	0.0093	0.0017
DCG score	α_4	-0.3668	0.0859
$E(\ln Y_i Y_i > 0)$			
Intercept	γ_0	7.6503	0.0909
MOVE! enrollment	γ_1	0.2153	0.0138
BMI	γ_2	0.0136	0.0018
Age	γ_3	-0.0186	0.0008
DCG score	γ_4	1.3492	0.0421
	ω	1.4123	0.0183
	κ	0.8428	0.0484

LSN, log-skew-normal; BMI, body mass index; DCG, diagnostic cost group.

These MTP model results suggest that VA expenditures are not reduced in the year following MOVE! initiation, possibly because few veterans have sustained an intense participation in this behavioral weight loss program [20]. This finding has important implications for VA policymakers needing to address the increasing incidence and prevalence of obesity among veterans. In particular, these results suggest that VA may need to introduce alternative weight management strategies to reduce expenditures among obese veterans or increase the effectiveness of MOVE! to induce expenditure reductions. It is possible that veterans' more recent (2012–2014) experience with MOVE! is more sustained and translates into VA expenditure reductions, but these findings suggest that enrollment in MOVE! in its initial 4 years was not associated with lower VA expenditures in the year following initiation, compared with non-enrollees.

5. Conclusion

We proposed an MTP model for semicontinuous data that allows investigators to obtain the population-average effect of covariates in the model. Our model directly parameterizes the covariates in terms of the population mean while still appropriately accounting for the excess number of zeros. While log-normal models are most commonly used, we also extended our MTP model to the broader and more flexible LSN distribution that allows asymmetry and skewness in the data and contains the log-normal distribution as a special case. In our simulation study, the maximum likelihood parameter estimates had near-zero bias and good coverage properties in all scenarios except for the intercept from the log-normal model when fit to skewed data. As such, using either the log-normal or LSN MTP model should be reliable for estimating effects of covariates, although additional care should be used when predicting marginal means for specified covariate groups.

Using the LSN MTP model, we estimated that enrollment in the MOVE! weight loss intervention was associated with 20% higher average health-care expenditures in the year after MOVE! initiation compared with a control group of non-enrollees. VA policymakers may need to refine the MOVE! program or introduce alternative behavioral weight programs to reduce expenditures among obese veterans. In contrast, the conventional two-part model found that MOVE! enrollment was associated not only with a decrease in the probability of incurring positive expenditures but also with an increase in the level of expenditures given they were incurred, leading to conflicting conclusions about the overall effect of the MOVE! program. Future directions for the MTP model could include extensions to clustered or spatially correlated data and applications to other fields, such as substance abuse or psychometric research. We are currently working to extend the MTP model to longitudinal data, allowing investigators to examine trends in the marginal mean over time. For example, it is possible that the effect of the MOVE! weight loss intervention may vary after additional years of enrollment, and estimating this effect would have important policy implications.

In short, the proposed MTP model provides a straightforward method for estimating covariate effects on the marginal mean of the population as a whole, which is not possible in conventional two-part models in a straightforward way. As such, it simplifies economic evaluations that are increasingly critical to understanding the return on investment of new interventions, policies, and programs.

Appendix: SAS code

A.1. Marginalized log-normal model

```
proc nlmixed data=mydata;
  bounds 0 <= sigma2;
  parms /* initial values for parameters */ ;
  linbin = a0 + a1*x1 + a2*x2;
  binprob = exp(linbin)/(1+exp(linbin)); /* probability y > 0 */
  mu = b0 + b1*x1 + b2*x2 - log(binprob) - sigma2/2;
  if y=0 then loglik=log(1-binprob);
  else if y>0 then loglik=log(binprob)-log(y)-.5*log
    (2*CONSTANT('PI'))-log(sqrt(sigma2))-(1/(2*sigma2))*(log(y)
    -mu)**2;
  model y~general(loglik);
  estimate 'marginal mean at x1=50 and x2=1' exp(b0+b1*50+b2*1);
run;
```

A.2. Marginalized log-skew-normal model

```
proc nlmixed data=mydata;
  bounds 0<=omega;
  parms /* initial values for parameters */ ;
  linbin = a0 + a1*x1 + a2*x2;
  binprob = exp(linbin)/(1+exp(linbin)); /* probability y > 0 */
  delta = kappa/sqrt(1+kappa**2);
  xi = b0 + b1*x1 + b2*x2 - log(2) - log(binprob) -
  log(CDF('NORMAL', omega*delta, 0, 1)) - (omega**2)/2;
  if y=0 then loglik=log(1-binprob);
  else if y>0 then do;
    pdfnormvar=(log(y)-xi)/omega;
    cdfnormvar=kappa*((log(y)-xi)/omega);
    loglik=log(binprob)+log(2)-log(y)-log(omega)+
    log(PDF('NORMAL', pdfnormvar, 0, 1)) +
    log(CDF('NORMAL', cdfnormvar, 0,1));
  end;
  model y~general(loglik);
  estimate 'marginal mean at x1=50 and x2=1' exp(b0+b1*50+b2*1);
run;
```

Acknowledgements

This work was funded by the Diabetes QUERI and the National Center for Health Promotion and Disease Prevention within the Department of VA. This work was also supported by the Office of Research and Development, Health Services Research and Development Service, Department of Veterans Affairs. Dr. Maciejewski is supported by a Research Career Scientist award (RCS 10-391) and a grant (IIR 10-159) from the VA. The Diabetes QUERI, the National Center for Health Promotion and Disease Prevention, and the Health Services Research and Development Service, Department of Veterans Affairs had no role in the design, conduct, collection, management, analysis, or interpretation of the data; or in the preparation, review, or approval of the manuscript. Dr. Maciejewski has received consultation funds from the Daiichi Sankyo and ResDAC at the University of Minnesota, and owns stock in Amgen due to his spouse's employment. No other authors have conflicts of interest.

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veteran Affairs, Duke University, or the University of North Carolina.

References

1. Manning WG, Morris CN, Newhouse JP, Orr LL, Duan N, Keeler EB, Leibowitz A, Marquis KH, Marquis MS, Phelps CE. A two-part model of the demand for medical care: preliminary results from the health insurance study. *Health, Economics, and Health Economics* 1981;103–123.
2. Duan N, Manning WG, Jr., Morris CN, Newhouse JP. A comparison of alternative models for the demand of medical care. *Journal of Business and Economic Statistics* 1983; 1(2):115–126. <http://www.jstor.org/stable/1391852>.
3. Xie H, McHugo G, Sengupta A, Clark R, Drake R. A method for analyzing longitudinal outcomes with many zeros. *Mental Health Services Research* 2004; 6:239–246.
4. Su L, Tom BDM, Farewell VT. Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics* 2009; 10(2):374–389.
5. Olsen MK, Schafer JL. A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association* 2001; 96(454):730–745.
6. Liu L, Strawderman RL, Johnson BA, O'Quigley JM. Analyzing repeated measures semi-continuous data, with application to an alcohol dependence study. *Statistical Methods in Medical Research* 2012. DOI: 10.1177/0962280212443324.
7. Aitchison J. On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association* 1955; 50(271):901–908.
8. Cragg JG. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 1971; 39(5):829–844. <http://www.jstor.org/stable/1909582>.
9. Azzalini A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 1985; 12:171–178.
10. Chai HS, Bailey KR. Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero. *Statistics in Medicine* 2008; 27(18):3643–3655.
11. Tooze JA, Grunwald GK, Jones RH. Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research* 2002; 11(4):341–355.
12. Mullahy J. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* 1998; 17:247–281.

13. Maciejewski ML, Liu CF, Kavee AL, Olsen MK. How price responsive is the demand for specialty care? *Health Economics* 2012; **21**:902–912.
14. Maciejewski ML, Smith VA, Livingston EH, Kavee AL, Kahwati LC, Henderson WG, Arterburn DE. Health care utilization and expenditure changes associated with bariatric surgery. *Medical Care* 2010; **48**(11):989–998.
15. Maciejewski ML, Livingston EH, Smith VA, Kahwati LC, Henderson WG, Arterburn DE. Health expenditures among high-risk patients after gastric bypass and matched controls. *Archives of Surgery* 2012; **147**(7):633–640.
16. Buntin MB, Zaslavsky AM. Too much ado about two-part models and transformation?: comparing methods of modeling medicare expenditures. *Journal of Health Economics* 2004; **23**(3):525–542.
17. Hernan MA, Robins JM. Causal inference, 2014. <http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>, Retrieved February 21, 2014 from Harvard University website.
18. Team RC. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2012. <http://www.R-project.org/>, ISBN3-900051-07-0.
19. Azzalini A. *R Package sn: The Skew-normal and Skew-t Distributions (Version 0.4-18)*. Università di Padova: Italia, 2013. <http://azzalini.stat.unipd.it/SN>.
20. Kahwati LC, Lance TX, Jones KR, Kinsinger LS. RE-AIM evaluation of the veterans health administration's MOVE! weight management program. *Translational Behavioral Medicine* 2011; **1**(4):551–560.
21. Arterburn DE, Maciejewski ML, Tsevat J. Impact of morbid obesity on medical expenditures in adults. *International Journal of Obesity Related Metabolic Disorders* 2005; **29**(3):334–339.
22. Finkelstein A, Trogdon JG, Cohen JW, Dietz W. Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health Affairs* 2009; **28**(5):w822–w831.
23. Berndt ER, Bir A, Busch SH, Frank RG, Normand SL. The medical treatment of depression, 1991–1996: productive inefficiency, expected outcome variations, and price indexes. *Journal of Health Economics* 2002; **21**(3):373–396.