ELSEVIER

Journal of Clinical Epidemiology 62 (2009) 953–958

Journal of Clinical Epidemiology

# Longitudinal tobit regression: A new approach to analyze outcome variables with floor or ceiling effects

Jos Twisk[a,b,*], Frank Rijmen[a]

[a]Department of Clinical Epidemiology and Biostatistics, VU Medical Centre, Amsterdam, The Netherlands
[b]Department of Methodology and Applied Biostatistics, Institute of Health Sciences, Vrije Universiteit, Amsterdam, The Netherlands

Accepted 1 October 2008

**Abstract**

**Background:** In many epidemiologic longitudinal studies, the outcome variable has floor or ceiling effects. Although it is not correct, these variables are often treated as normally distributed continuous variables.

**Objectives:** In this article, the performance of a relatively new statistical technique, longitudinal tobit analysis, is compared with a classical longitudinal data analysis technique (i.e., linear mixed models).

**Study Design and Setting:** The analyses are performed on an example data set from rehabilitation research in which the outcome variable of interest (the Barthel index measured at on average 16.3 times) has typical floor and ceiling effects. For both the longitudinal tobit analysis and the linear mixed models an analysis with both a random intercept and a random slope were performed.

**Results:** Based on model fit parameters, plots of the residuals and the mean of the squared residuals, the longitudinal tobit analysis with both a random intercept and a random slope performed best. In the tobit models, the estimation of the development over time revealed a steeper development compared with the linear mixed models.

**Conclusion:** Although there are some computational difficulties, longitudinal tobit analysis provides a very nice solution for the longitudinal analysis of outcome variables with floor or ceiling effects. 

*Keywords:* Longitudinal studies; tobit analysis; Linear mixed models; Statistical methods; Floor effects; Ceiling effects

## 1. Introduction

Within epidemiology, there is an increasing interest in performing prospective cohort studies. One of the purposes of these studies is to analyze the longitudinal development over time in a particular outcome variable. In some of these studies, the outcome variable of interest reaches a certain ceiling over time. For instance, in rehabilitation research, most of the patients will recover after a certain amount of time. On the instrument to measure the rehabilitation process, these patients cannot score any higher than the maximum. It is also possible that so-called floor effects occur. For instance, when pain medication is investigated and the outcome variable pain is measured on a visual analog scale, some patients will report ''no pain'' after a certain amount of time. They cannot score lower than the ''no pain'' level. Also in studies where there is some detection limit (e.g., for blood parameters or environmental factors, such as pesticides), these floor effects are present. In fact these problems always arise when a measurement instrument that has upper and lower limits is used and when some of the patients in the study reach these upper or lower limits. One can think of an underlying latent variable with an unrestricted range, of which the observed outcome is an on both sides truncated version, so that floor and ceiling effects can be considered to be a kind of interval censoring. Sometimes, floor and ceiling effects are referred to as lower and upper censoring.

In most longitudinal epidemiological studies, these floor and ceiling effects are ignored. The development over time of such outcome variables are analyzed as if they were normally distributed over the whole period of time. This is not the case, because when patients reach the floor or ceiling, the outcome variable is not normally distributed anymore. In cross-sectional studies (especially in econometrics), the problem of upper and lower censoring is solved by using so-called tobit models, after Tobin's [1] classical example on household expenditures. Within epidemiology, only a few examples are available in which cross-sectional tobit

* Corresponding author. Department of Clinical Epidemiology and Biostatistics, VU Medical Centre, de Boelelaan 1118, 1081 HV Amsterdam, The Netherlands.
*E-mail address*: jwr.twisk@vumc.nl (J. Twisk).

0895-4356/09/$ – see front matter 
doi: 10.1016/j.jclinepi.2008.10.003

**What is New?**

Longitudinal tobit analysis is suitable for the analysis of longitudinal data with floor and/or ceiling effects and it outperforms the traditional linear mixed models.

Although there could be some computational difficulties, longitudinal tobit analysis can be performed within Stata software.

analysis is used [2–9]. However, for longitudinal epidemiological studies, tobit analysis is (to our knowledge) never used, although it has some nice theoretical advantages above the "classical" longitudinal data analysis.

The purpose of this study is to compare longitudinal tobit analysis with "classical" longitudinal analysis to investigate the longitudinal development over time in outcome variables with floor and/or ceiling effects. The example used in the present article is taken from rehabilitation research.

## 2. Methods

### 2.1. Study population

The population used in the present study was taken from a longitudinal rehabilitation study among stroke patients [10]. The main purpose of the study was to analyze the development of the Barthel index. An outcome variable that represents a patient's ability to carry out 10 everyday tasks (i.e., bladder and bowel control, toilet use, dressing, feeding, walking, personal toilet, transfer activities, bathing, and stair climbing) [11]. The lowest score for the Barthel index is 0 and the highest possible score is 20. The Barthel index was assessed weekly during the first 10 weeks after stroke onset, then every 2 weeks until week 20 and finally the Barthel index was assessed at week 26, week 38, and week 52. The study population consisted of 101 patients with on average 16.3 measurements (range, 2–18) per patient. Forty-seven patients had a full data set, whereas 33 patients only missed the first measurement. Furthermore, there were seven patients who dropped-out (varying after the second measurement to the measurement after 26 weeks), there were six patients with intermittent missing data with more than three missing observations, and eight patients with intermittent missing data with only one or two missing observations.

Besides the outcome variable, several covariates were measured at baseline: sit-balance, incontinence, type of stroke, and age. For detailed information see Kwakkel et al. [10].

In the example, two research questions will be addressed. First, the longitudinal development over time will be analyzed, and second, the influence of the covariates measured at baseline will be analyzed. The development over time was modeled as a second order polynomial function.

### 2.2. Statistical analysis

#### 2.2.1. Classical longitudinal analysis

The classical statistical methods to be used to answer the above research questions are either linear mixed models [12], which are also known as multilevel models, hierarchical models, or random coefficient models or generalized estimating equations (GEE analysis) [13]. In the present example, only linear mixed models will be used because for continuous outcome variables, linear mixed models are, in general, a bit more flexible compared with GEE analysis [14]. Two different analyses will be presented. First, an analysis with only a random intercept and second, an analysis with both a random intercept and a random slope for time.

#### 2.2.2. Tobit longitudinal analysis

The general idea of tobit regression is that it models both the probability of reaching either the floor or ceiling and the development over time between the floor and ceiling. The tobit model originated in the context of linear regression analysis (cross-sectional data), and can be formulated mathematically as follows. Let $y^*$ be a random latent variable that is not censored, and assume a linear regression model for it:

$$y_i^* = x_i'\beta + e_i, \quad e_i \sim N(0, \sigma^2)$$

where $i$ refers to subject $i$.

Furthermore, it is assumed that we can observe the realizations of $y^*$ for a given range $[l,u]$ only, and that values of $y^*$ smaller than $l$ or larger than $u$ are censored at, respectively, $l$ and $u$. Hence, the observed limited dependent variable $y$ is obtained from $y^*$ as

$$\begin{aligned} y_i &= l \quad \text{for } y_i^* \leqslant l \\ y_i &= y_i^* \quad \text{for } l < y_i^* < u \\ y_i &= u \quad \text{for } y_i^* \geqslant u \end{aligned}$$

If a dependent variable is limited at one side, only a lower (or upper) limit is needed ($l = -\text{inf}$ or $u = +\text{inf}$).

Because of the censoring mechanism, $E(y)$ is not equal to $E(y^*)$. Because the distribution for $y$ is not the same as the distribution for $y^*$, the expected values will be different. Therefore, parameter estimates may become inconsistent.

For longitudinal data, a tobit model can be defined in a similar way. As in classical longitudinal analysis, a natural choice for the underlying model for $y^*$ is the linear mixed model [12]:

$$\begin{aligned} &y_{ij}^* | b_i = \mathbf{x}_{ij}'\boldsymbol{\beta} + z_{ij}'\mathbf{b}_i + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2) \\ &\mathbf{b}_i \sim N(0, \mathbf{D}) \end{aligned}$$

where $i$ refers to case $i$ and $j$ to the $j$th measurement. That is, conditional on the case-specific parameters $\mathbf{b}_i$, a linear

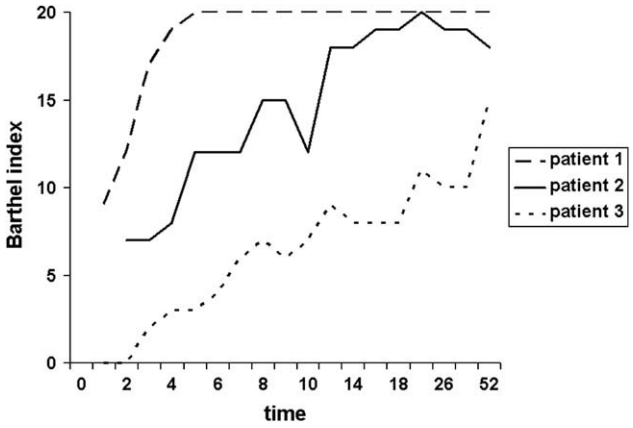


Fig. 1. Development of the Barthel index over time for three subjects in the study population.

model is assumed with $E(y^*_{ij}|\mathbf{b}_i)=\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{b}_i y$ is obtained from $y^*$ as before.

All analyses were performed with Stata [15]. Estimating the parameters of a longitudinal tobit model is complicated by the fact that the likelihood contains a not analytically solvable integral over the random effects $\mathbf{b}_i$. When the dimensionality of $\mathbf{b}_i$ is low, the integral can be approximated using Gaussian quadrature, an approach that is implemented in the *GLLAMM* procedure [16,17]. Tobit regression with only a random intercept, however, is less complicated and can be estimated with the *xttobit* procedure. ''Classical'' linear mixed model analysis was performed with the *xtmixed* procedure.

To compare the longitudinal linear mixed model analysis with the tobit longitudinal analysis, first Akaike's information criterion (AIC) and Bayesian information criterion (BIC) were used. Both are indicators of model fit, taking into account the number of parameters to be estimated in the different models. Second, residual plots were made and the means of the squared residuals were computed for the different models. This is to compare the precision of the prediction based on the estimated regression coefficients of the different models. The residuals are calculated as the difference between the observed values and the predicted values. These predicted values include the random effects and can be larger than 20 or lower than 0.

## 3. Results

Figure 1 shows three typical examples of patients in the present study. The first patient reaches the maximum score after 5 weeks, whereas the second patient starts with a Barthel index of 7, increases to the highest level at 20 weeks, and decreases again to a score of 18 at week 52. The third patient starts at the lowest level, stays at the lowest level for 2 weeks, and never reaches the maximum score.

Table 1 shows the results of the analysis to investigate the longitudinal development over time. Based on the AIC and BIC, it is obvious that the longitudinal tobit model performs much better than the linear mixed model. The regression coefficient for the linear time component is lower in the linear mixed model analysis compared with the tobit analysis, whereas the quadratic time component is estimated at about the same value. This means that regression curves as a function of time are steeper for the longitudinal tobit model than for the linear mixed model.

The results of the fit statistics are more or less confirmed by the histograms of the residuals of the four different models (Fig. 2). To illustrate the specific pattern of the residuals as a function of time, Fig. 3 shows this pattern for the linear mixed model with both a random intercept and a random slope. It can be seen that the outcome variable is overestimated in the first measurements, at the intermediate measurements the outcome variable is underestimated, whereas at the last measurements the outcome variable is overestimated again. This U-shaped pattern is present in both the linear mixed models and the longitudinal tobit analysis, but in the latter, the absolute values of the over- and underestimations are much lower (data not shown for all four methods). Surprisingly, at the last measurement at 52 weeks, the outcome variable is underestimated again. This phenomenon is probably caused by the quadratic nature of the longitudinal development.

The mean squared residuals (see Table 2), show a slightly different pattern than the model fit parameters. In fact,

Table 1
Regression coefficients (and standard errors) and model fit indices for four different analyses to analyze the longitudinal development over time for the Barthel index

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| Time | 0.664 (0.015) | 0.664 (0.016) | 0.766 (0.017) | 0.855 (0.042) |
| Time$^2$ | −0.010 (0.0003) | −0.010 (0.0003) | −0.011 (0003) | −0.010 (0.0003) |
| AIC | 7864 | 7711 | 7035 | 6727 |
| BIC | 7891 | 7743 | 7062 | 6765 |
| Mean squared residual (standard deviation) | 5.03 (7.87) | 3.96 (6.46) | 4.49 (9.28) | 3.03 (5.14) |

Model 1: Linear mixed model analysis with a random intercept.
Model 2: Linear mixed model analysis with a random intercept and a random slope for time.
Model 3: Longitudinal tobit analysis with a random intercept.
Model 4: Longitudinal tobit analysis with a random intercept and a random slope for time.

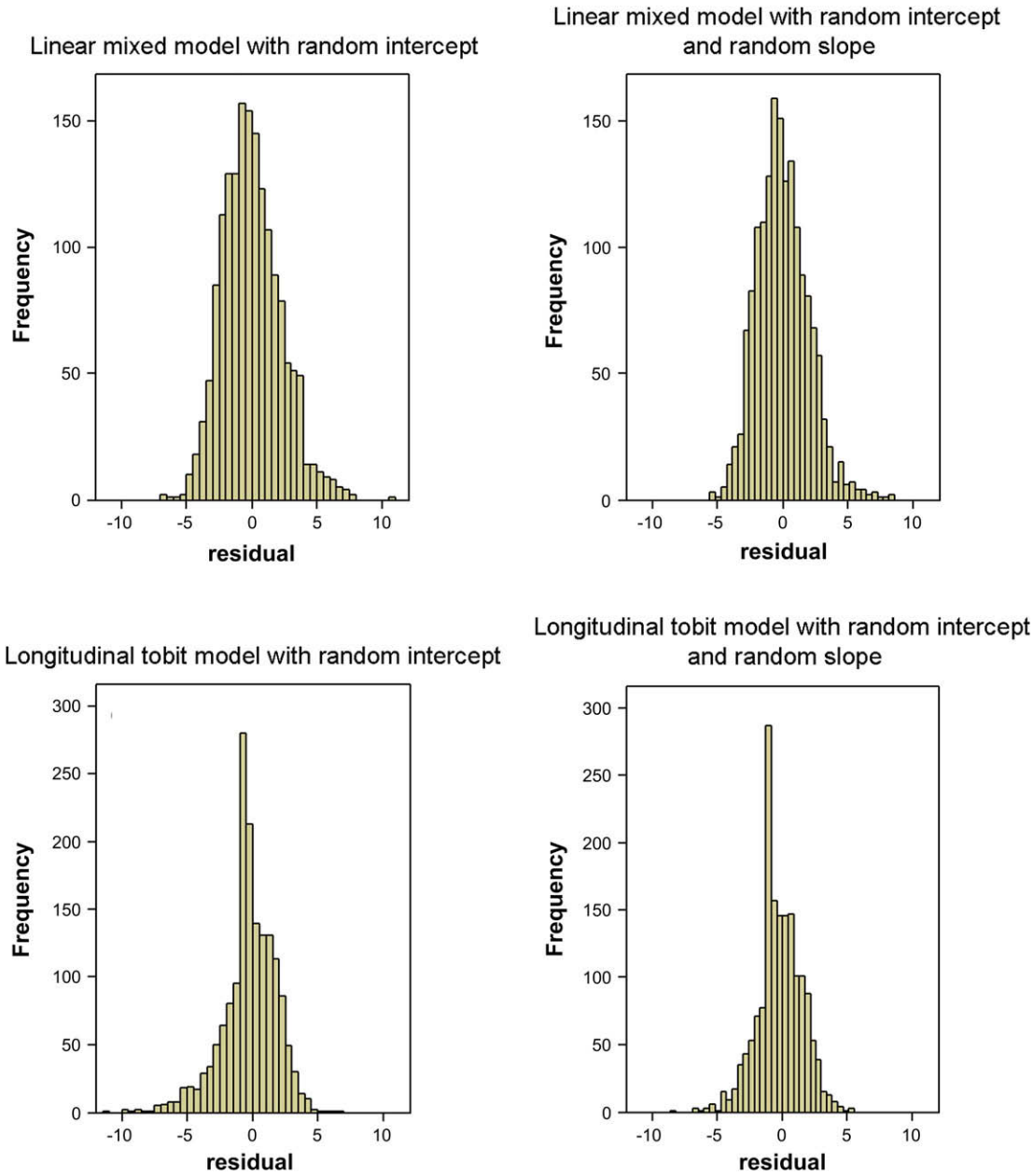


Fig. 2. Histograms of the residuals of the four different models without covariates.

estimating a random slope for time seems to be slightly more important than taking into account the ceiling effect. However, also regarding the mean squared residuals, tobit analysis with both a random intercept and a random slope seems to perform best.

Table 2 shows the results of the analysis to investigate the influence of certain covariates on the longitudinal development over time in the Barthel index. Based on the AIC and BIC values, the longitudinal tobit models are again to be preferred to the linear mixed models. This finding is further supported by smaller means of the squared residuals under the longitudinal tobit models. The scatterplots of the residuals show exactly the same pattern as for the analysis without covariates; that is, the estimated increase in the Barthel index as a function of time was higher for the longitudinal tobit model than for the linear mixed model (data not shown). Regarding the magnitude of the regression coefficients for the covariates, no clear pattern was found.

## 4. Discussion

The results of the present study show that the longitudinal tobit models give better model fits compared with the linear mixed models, both with and without taking covariates into account. The improvement over time was estimated to be faster in the longitudinal tobit model than in the linear mixed model. This is not very strange because

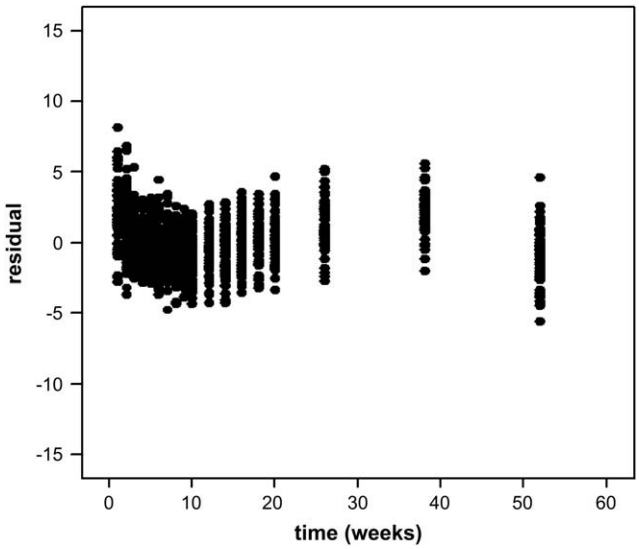


Fig. 3. Scatterplot of the residuals as a function of time for the linear mixed model with both a random intercept and slope without covariates.

due to the substantial upper censoring present in the current data the regression line (estimated with the linear mixed models) is pulled down [18]. This phenomenon also explains the specific pattern in the scatterplots of the residuals by time. By ''pulling down'' the regression line for time, the values of the Barthel index get underestimated. At later measurement occasions, this effect is counterbalanced by the censoring mechanism: due to the improvement over time, the linear mixed model increasingly estimates the values of the Barthel index at values above 20, the upper limit, and hence the values of the Barthel index are overestimated.

In this example, a measurement instrument is used with a fixed ceiling. The Barthel index cannot go higher than 20 for all patients. In growth research, however, there are also examples that there is a certain limit for each patient, but that that limit is not equal for all patients. Body height is probably the best example of such an outcome variable. The alternative way to analyze the longitudinal development of such outcome variables is called interval regression. This technique is based on more or less the same methodological background as tobit analyses, but it goes one step further [18] and it allows upper and lower censoring at subject-specific values.

Sometimes the problem of lower censoring is tackled by using zero-inflated Poisson models [19–21]. However, these models cannot be used in the general situation with lower censoring, because the underlying distribution of the outcome variable has to be Poisson. In fact, zero-inflated Poisson models are specifically developed for ''count'' data suffering from too many zeros.

As mentioned earlier, tobit regression analysis originated in econometrics and it is not much used in epidemiology, although in many epidemiological studies outcome variables that are either upper or lower censored are used. There are only a few examples in the literature and all these studies have a cross-sectional design. For instance, Xu et al. [6] used tobit analysis to investigate health-related quality of life of patients after hip arthroplasty, whereas Delva et al. [8] used tobit analysis to study the relationship between alcohol abuse and depression among adolescent females. Finally, Ferraro et al. [2] investigated the relationship between body mass index and disability by using tobit analysis.

### 4.1. What method should be used?

Theoretically, the longitudinal tobit analyses are better suited for the analyses of the longitudinal development of outcome variables that are either upper or lower censored. This is also confirmed by the better model fits for these

Table 2
Regression coefficients (and standard errors) and model fit indices for four different analyses to analyze the relationship between baseline covariates and the longitudinal development over time of the Barthel index

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| Time | 0.663 (0.015) | 0.664 (0.016) | 0.765 (0.017) | 0.851 (0.042) |
| Time$^2$ | −0.010 (0.003) | −0.010 (0.0003) | −0.011 (0.0003) | −0.010 (0.0003) |
| Age | −0.118 (0.030) | −0.103 (0.030) | −0.143 (0.033) | −0.039 (0.030) |
| Sit-balance | 3.919 (0.826) | 3.580 (0.833) | 4.124 (0.923 | 3.263 (0.745) |
| Incontinence | −0.704 (0.758) | −1.504 (0.765) | −0.862 (0.848) | −1.828 (0.687) |
| Type 1[a] | −3.735 (1.280) | 4.567 (1.293) | −5.250 (1.442) | −3.409 (1.257) |
| Type 2 | −5.231 (1.344) | −5.877 (1.358) | −6.915 (1.514) | −3.975 (1.340) |
| AIC | 7794 | 7637 | 6963 | 6679 |
| BIC | 7848 | 7697 | 7017 | 6744 |
| Mean squared residual (standard deviation) | 5.03 (7.82) | 3.96 (6.44) | 4.49 (9.19) | 3.02 (5.08) |

Model 1: Linear mixed model analysis with a random intercept.
Model 2: Linear mixed model analysis with a random intercept and a random slope for time.
Model 3: Longitudinal tobit analysis with a random intercept.
Model 4: Longitudinal tobit analysis with a random intercept and a random slope for time.
[a] Type of stroke: type 1 is a partial anterior cerebral infarct (PACI), type 2 is a total anterior cerebral infarct (TACI) and the reference type is a lacunar circulation infarct.

models compared with the linear mixed models. However, it should be borne in mind that the parameters of the longitudinal tobit analysis are difficult to estimate, especially when there are more random coefficients and/or the model becomes more extensive. So, because of the computational difficulty and instability of the models, one should be very careful with the use of complicated longitudinal tobit models. Furthermore, the smaller the number of censored observations, the more the likelihood of the longitudinal tobit model will resemble that of the linear mixed model. Hence, when the number of censored observations constitutes only a small proportion of the data, the linear mixed model will produce estimates (and standard errors) that are biased to a small degree only, and may be, therefore, preferred.

## Appendix

### Likelihood for the mixed tobit model

The mixed tobit model is defined through a linear mixed model on underlying latent variables $y^*_{ij}$,

$$y^*_{ij}|b_i = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2)$$
$$\mathbf{b}_i \sim N(0, \mathbf{D}),$$

where $i$ refers to case $i$ and $j$ to the $j$th measurement.

Conditional on the case-specific parameters $\mathbf{b}_i$, a linear model is assumed with $E(y^*_{ij}|\mathbf{b}_i) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$, where $y$ is obtained from $y^*$ as:

$$y_{ij} = l \quad \text{for } y^*_{ij} \leqslant l$$
$$y_{ij} = y^*_{ij} \quad \text{for } l < y^*_{ij} < u$$
$$y_{ij} = u \quad \text{for } y^*_{ij} \geqslant u$$

Hence, the density function of $y$ is:

$$f(y_{ij} = l) = F\left(y^*_{ij} = l\right)$$
$$f(y_{ij}) = f\left(y^*_{ij}\right) \quad \text{for} l < y_{ij} < u$$
$$f(y_{ij} = u) = 1 - F\left(y^*_{ij} = u\right)$$

The contribution to the likelihood of a case $i$ is obtained as the product over the $J$ measurements for case $i$, and integrating this product over the case-specific parameters $\mathbf{b}_i$,

$$L_i = \int_{\mathbf{b}_i} \prod_j f(y_{ij}) N(\mathbf{b}_i; 0, \mathbf{D}) d\mathbf{b}_i.$$

Finally, the likelihood for all cases is:

$$L = \prod_i L_i$$

## References


[1] Tobin J. Estimation of relationships for limited dependent variables. Econometrics 1958;26:24–36.

[2] Ferraro KF, Su Y, Gretebeck RJ, Black DR, Badylak SF. Body mass index and disability in adulthood: a 20-year panel study. Am J Public Health 2002;92:834–40.

[3] Marshall TA, Levy SM, Broffitt B, Warren JJ, Eichenberger-Gilmore JM, Burns TL, Stumbo PJ. Dental caries and beverage consumption in young children. Pediatrics 2003;112: 184–91.

[4] Gasser T, Rousson V. Modelling neuromotor ratings with floor-effects. Stat Med 2004;23:3641–53.

[5] Lubin JH, Colt JS, Camann D, Cavis S, Cerhan JR, Severson RK, Bernstein L, Hartge P. Epidemiologic evaluation of measurement data in the presence of detection limits. Environ Health Perspect 2004;112:1691–6.

[6] Xu M, Garbuz DS, Kuramoto L, Sobolev B. Classifying health-related quality of life outcomes of total hip arthroplasty. BMC Musculoskelet Disord 2005;6(48):1–9.

[7] Brennan DS, Spencer AJ. Mapping oral health related quality of life to generic health state values. BMC Health Serv Res 2006;6(96): 1–10.

[8] Delva J, Grogan-Kaylor A, Steinhoff E, Shin D-E, Siefert K. Using tobit regression analysis to further understand the association of youth alcohol problems with depression and parental factors among Korean adolescent females. J Prev Med Public Health 2007;40: 145–9.

[9] Scirica CV, Gold DR, Ryan L, Albukerim H, Celedon JC, Platts-Mills TAE, Naccara LM, Weiss ST, Litonjua AA. Predictors of cord blood IgE levels in children with risk for asthma and atopy. J Allergy Clin Immunol 2007;119:81–8.

[10] Kwakkel G, Wagenaar RC, Twisk JWR, Lankhorst GJ, Koetsier JC. Intensity of leg and arm training after primary middle-cerebral-artery stroke: a randomised trial. Lancet 1999;354:191–6.

[11] Wade DT, Collin C. The Barthel ADL index: a standard measure of physical disability? Int Disabil Stud 1988;10:64–7.

[12] Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer-Verlag; 2000.

[13] Zeger SL, Liang K-Y. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30.

[14] Twisk JWR. Longitudinal data analysis. A comparison between generalized estimating equations and random coefficient analysis. Eur J Epidemiol 2004;19:769–76.

[15] StataCorp. Stata Statistical Software, Release 9.0 College Station, Texas: StataCorp; 2006.

[16] Rabe-Hesketh S, Pickles A, Skondral A. GLLAMM manual. U.C. Berkeley Division of biostatistics working paper series. Working paper 160. Available at http://www.bepress.com/ucbbiostat/paper160; 2004.

[17] Rabe-Hesketh S, Skondral A. Multilevel and longitudinal modeling using Stata. College Station, Texas: Stata Press; 2005.

[18] Long JS. Regression models for categorical and limited dependent variables. Thousand Oaks, California: Sage Publications; 1997.

[19] Cheung YB. Zero-inflated models for regression analysis of count data: a study of growth and development. Stat Med 2002;21:1461–9.

[20] Slymen DJ, Ayala GX, Arredondo EM, Elder JP. A demonstration of modelling count data with an application to physical activity. Epidemiol Perspect Innovat 2006;3(3):1–9.

[21] Horton NJ, Kim E, Saitz R. A cautionary note regarding count models of alcohol consumption in randomised controlled trials. BMC Med Res Method 2007;7(9):1–9.