

Analysing outcome variables with floor effects due to censoring: a simulation study with longitudinal trial data

Alette S. Spiensma⁽¹⁾, Iris Eekhout^(1,3), Michiel R. de Boer⁽²⁾, Jolanda J. Luime⁽⁶⁾, Pascal H. de Jong⁽⁵⁾, Melike Kaya Bahçecitapar⁽⁴⁾, Martijn W. Heymans⁽¹⁾, Jos W.R. Twisk⁽¹⁾

(1) VU University Medical Center, Department of Epidemiology and Biostatistics, Amsterdam Public Health research institute, Amsterdam, The Netherlands

(2) VU University, Department of Methodology and Applied Biostatistics, Faculty of Earth and Life Sciences, Institute of Health Sciences, Amsterdam, The Netherlands

(3) Netherlands Organisation for Applied Scientific Research TNO, Leiden, The Netherlands

(4) Hacettepe University, Department of Statistics, Faculty of Science, Ankara, Turkey

(5) Erasmus University Medical Center, Department of Rheumatology, Rotterdam, The Netherlands

CORRESPONDING AUTHOR: JWR Twisk, VU University medical center, Department of Epidemiology and Biostatistics, Medical Faculty (F-wing), De Boelelaan 1089a, 1081 HV Amsterdam, The Netherlands

DOI: 10.2427/12850

Accepted on May 22, 2018

ABSTRACT

Background: Randomised controlled trials (RCTs) are the gold standard to estimate treatment effects. When patients receive effective treatment over time they may reach the limit of a certain measurement scale. This phenomenon is known as censoring and leads to skewed distributions of the outcome variable with an excess of either low (floor effect) or high values (ceiling effect). Applying traditional methods such as linear mixed models to analyse these kind of longitudinal RCT data may result in bias of the regression coefficients. To deal with floor effects due to censoring, a tobit mixed model can be used. The objective of this study was to compare the results of longitudinal linear mixed model analyses with longitudinal tobit mixed model analyses.

Methods: A simulation study was performed in which several situations of RCTs with floor effects were simulated. From the simulated datasets, which were set up to estimate the interaction between treatment and time, the regression coefficient for this interaction and for the overall treatment effect were estimated. Additionally, data from an empirical RCT were analysed with both methods.

Results: Regarding the interaction between treatment and time, the results of the tobit mixed model analysis were the same as the true values in all conditions, while the linear mixed model analysis revealed highly underestimated regression coefficients. However, the overall treatment effect with an increasing number of follow-up measurements in combination with a strong floor effect showed that the estimates from the tobit mixed model were also not accurate.

Conclusion: Tobit mixed model analysis should be used to estimate treatments effects in longitudinal RCTs with floor effects due to censoring.

Key words: Tobit, Simulation, Longitudinal, Floor effects, Excess of zeroes, Epidemiology

INTRODUCTION

Randomized controlled trials (RCTs) are assumed to be the gold standard to estimate treatment effects [1]. Because RCTs often include multiple follow-up measurements, longitudinal statistical methods have to be used to estimate the treatment effects [2]. When patients receive effective treatment over time it may result in reaching the limit of a certain scale. When many patients reach this limit, it results in a skewed distribution of the outcome with an excess of either the lowest (floor effect) or the highest score (ceiling effect) on the scale of measurement. In some instances the true outcome follows a normal distribution, but values below a certain threshold are not detected, a phenomenon which is also known as censoring [3]. In epidemiological and clinical studies, there are many examples of outcome measures that show floor or ceiling effects over time due to censoring. Functional ability measures, such as the Disability Index of the Health Assessment Questionnaire (HAQ-DI) (floor effect) and the Barthel index (ceiling effect), are especially prone to this phenomenon [4]. The problem with, for instance, the floor effect of HAQ-DI is that patients that score zero should not be regarded as patients that truly all have the same score. For some of the patients the true score may fall beyond the scale of the measurement instrument. In other words, there is a certain variance between patients at the limit that cannot be observed.

Reviewing the medical literature, it is striking to see how these kind of data (i.e. skewed to the right longitudinal RCT outcomes with floor effects due to censoring) are analysed. Non-parametric testing is mostly used which ignores the longitudinal nature of the data and does not lead to an effect estimate [5]. When the longitudinal nature of the data is not ignored, traditional longitudinal methods, such as linear mixed models or linear generalised estimation equations (GEE- analysis) are used [6,7]. However, these linear methods assume a normal distribution of the outcome. Previous research has shown that when a normal distribution is assumed while a (strong) floor effect is present, regression parameters become biased and artefactual non-linearity can occur [8-11]. To deal with either floor or ceiling effects, a tobit mixed model analysis can be used. The tobit mixed model assumes that there is a certain variation in the upper or lower limit of the scale [12]. The tobit mixed model [13,14] has rarely been used to analyse RCT data with floor effects in epidemiological and clinical studies. On the other hand, it is not clear to what extent longitudinal linear mixed models are able to handle floor effects due to censoring, and when it is necessary to use a longitudinal tobit model. Therefore, the objective of this study was to compare the performance of a longitudinal linear mixed model with a longitudinal tobit mixed model in different data situations of floor effects in a longitudinal RCT. We simulated data in which we varied the number of follow-up measurements, the sample size, the magnitude of the treatment effect and

the number of zeroes in the follow-up measurements of the outcome variable. Furthermore, both methods were used to analyse data from an empirical RCT.

METHODS

Simulation

With Stata [15], we simulated 'true' longitudinal RCT datasets with a normally distributed outcome variable. Table 1 shows the Stata syntax used to set up the simulations.

The simulation set up shown in Table 1 included 3 measurements (one baseline and two follow-up measurements) on 100 patients. A dichotomous treatment variable was created that randomly and evenly assigned the simulated patients to either the control group or the intervention group. The parameters used for the simulation set up were derived from the empirical dataset. Both the random intercept variance and the residual variance were about 0.5 (i.e. the square root of 0.7). The model which was used for the simulation included - besides the random intercept variance and the residual variance - a fixed intercept of 1.5, a regression coefficient for time of -0.1 and a regression coefficient for the time*treatment interaction of -0.2, leading to an outcome variable which was normally distributed. The inverse sign of the two regression coefficients indicate a linear decrease in outcome over time for the control group and a stronger linear decrease over time in outcome for the intervention group. In order to generate floor effects, all negative outcome values were set to zero. Table 2 shows the different conditions that were used in the simulations. A total of 12 conditions were compared and for each condition 1000 samples were generated.

Figure 1 shows the percentage of patients with floor effects for the different conditions, separately for the control and treatment group over time. The magnitude of the treatment effect and the number of follow-up measurements were related to the amount of zeroes. A stronger treatment effect and more follow-up measurements resulted in more patients reaching scores below zero on the scale of the outcome measurement.

Modelling simulated data with and without floor effects

For all analyses, we used mixed models with a random intercept to take into account the dependency of the observations within the patient in the simulated longitudinal data [6,16]. The following analyses were performed: 1) a linear mixed model analysis on the simulated RCT data without floor effects. 2) a linear mixed model analysis on the simulated RCT data with floor effects

and 3) a tobit mixed model analysis on the simulated RCT data with floor effects. In all models, treatment, time as well

as the interaction between treatment and time were added (see Table 1, Table 2 and Equation 1).

TABLE 1. Syntax to simulate RCT datasets with an excess of zeros

```
clear
program zero
drop _all
set obs 100 // 100 patients
gen pat_id=_n //patient id
generate nu0 = sqrt(0.7)*rnormal() //random intercept per patient
gen treatment= runiform() //treatment variable
replace treatment=0 if treatment <0.5
replace treatment=1 if treatment >0.5
expand 3 // 3 timepoints per patient
bysort pat_id: gen index = _n
generate time = index - 1
gen residual = sqrt(0.7)* rnormal() // residual
gen true_outcome = -.1*time + -.2*time*treatment + 1.5
gen outcome = true_outcome + nu0 + residual
replace outcome=0 if outcome<0 // generate floor effects*
mixed##c.time || pat_id: //
                                linear mixed model analysis
end
simulate _b _se, seed(12345) reps(1000): zero
```

*In the original simulations, negative values were not replaced by zero

TABLE 2. Different conditions used in the simulations

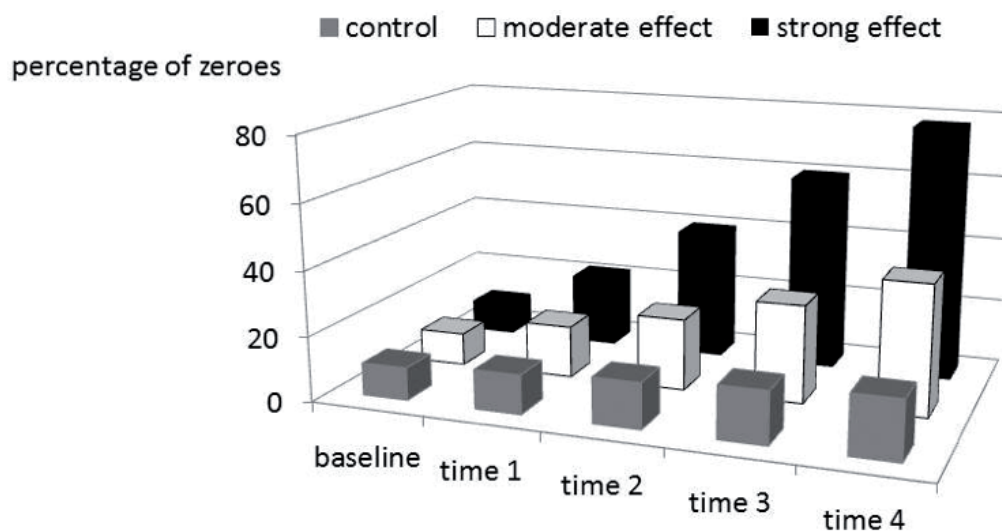
```
set obs 100 // 100 patients
set obs 200 // 200 patients
set obs 300 // 300 patients

expand 3 // 3 timepoints per patient
expand 5 // 5 timepoints per patient

gen true_outcome = -.1*time + -.2*time*treatment + 1.5 // moderate
                                treatment effect
gen true_outcome = -.1*time + -.5*time*treatment + 1.5 // strong
                                treatment effect

mixed outcome c.treatment##c.time || pat_id: //
                                linear mixed model analysis
xtset pat_id
xttobit outcome c.treatment##c.time // tobit mixed model analysis
```

FIGURE 1. Percentage of patients with floor effects in the censored simulated data for the different conditions in the control and treatment group at baseline and the follow-up measurements.



$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + \beta_2 x_{2i} + \beta_3 x_{3ij} + \epsilon_{ij} \quad (1)$$

where y_{1ij} is the outcome; $(\beta_0 + \mu_{0i})$ is the patient specific intercept, x_{1ij} is the covariate time, x_{2i} is the covariate treatment, x_{3ij} is the time treatment interaction covariate, and ϵ_{ij} is the time specific error component. The tobit mixed model is specified in a similar fashion, only the y_{ij} is replaced by y_{ij}^* , which indicates the difference between the linear mixed model and the tobit mixed model, i.e. for the latter the outcome is estimated on a latent (unobservable) scale, allowing some variation in the lower limit of the scale.

For both models the interaction between treatment and time is the most important parameter, because this reflects the difference in change over time between the intervention and control group, which can be considered as the treatment effect.

In a second analysis, from the same simulations with the same three methods, the overall difference between the treatment groups over time was estimated (see Equation 2).

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1i} + \epsilon_{ij} \quad (2)$$

where y_{1ij} is the outcome; $(\beta_0 + \mu_{0i})$ is the patient specific intercept, x_{1i} is the covariate treatment, and ϵ_{ij} is the time specific error component.

Where a linear mixed model analysis assumes a normal distribution for all patients, even though there might be a skewed distribution due to the high amount of zeroes, the tobit mixed model assumes a normal distribution for the patients without the floor effects and combines this with the assumption that the patients with floor effects are not all equal with respect to their observed outcome scores. Actually, the tobit model assumes that the observed outcomes result from a latent normally distributed variable, where observations are censored if they fall below a certain threshold. This enables the estimation of values outside the range of the observed patient scores, which may better match the 'true' variation in outcome scores of the patients [17].

Both linear mixed model analyses and tobit mixed model analyses were performed in Stata [15].

Performance of the models

Based on Burton et al. [18], we examined bias, standardized bias and coverage of regression coefficients to compare the performance of the linear mixed model and the tobit mixed model. Bias was determined by comparing the difference between the true regression coefficients of the interaction between treatment and time (i.e. -0.2 for the moderate treatment effect and -0.5 for the strong treatment effect) to the regression coefficients of the linear mixed model analysis and tobit mixed model analysis obtained

from the data with floor effects. For illustrative purposes we also calculated the regression coefficients for the simulated datasets without censoring.

Standardized bias was calculated by expressing the bias as a percentage of the standard error of the estimate. A value above 40 percent, both negative as well as positive, was considered poor [19]. The percentage of times that the confidence interval of the calculated regression coefficients included the 'true' regression coefficient was used to assess the coverage. Since we use a 95% confidence interval, the ideal coverage should have a score of 95%. Values above 95% (over-coverage) reflect a too high type II error rate in the parameter estimates, values under 95% will lead to an increase of type I errors in the parameter estimates[19]. All performance parameters were calculated in Stata [15].

RESULTS

Table 3 shows the regression coefficients and standard errors (SE) of the treatment and time interaction for all simulated datasets. It can be seen that the results of the tobit mixed model analysis were the same as the true values as well as the results obtained from the mixed model analysis on the datasets without floor effects. The results of the linear mixed model analysis on the datasets with floor effects showed high underestimations of the regression coefficients, which became stronger when the treatment effect and the number of follow-up measurements increased. All differences were independent of the sample size, The differences between the methods are also reflected in the standardised bias and coverage (see Table 4).

When the overall treatment effect was considered, the regression coefficients of the tobit mixed model analysis were slightly lower than the ones obtained from the linear mixed model analysis on the datasets without floor effects. Because the simulated datasets were not set up to estimate the overall treatment effect, we could not compare the results with the true values. For the linear mixed model analyses on the datasets with floor effects, the underestimation of the regression coefficients were much stronger than for the tobit mixed model analyses (see Table 5). Again the differences between the methods were stronger when percentage of zeroes became higher (i.e. stronger treatment effect and more follow-up measurements) and were also reflected in the standardised bias and coverage of the different analyses (see Table 6). According to the standardized bias and the coverage, in the dataset with the highest percentage of zeroes, the tobit mixed model analyses also did not perform well in estimating the overall treatment effect.

Empirical example

To illustrate the differences between a linear mixed

TABLE 3. Regression coefficients and standard errors (SE) of the treatment and time interaction estimated in different simulated datasets
2 follow-up measurements

	N	Moderate effect (-0.20)			Strong effect (-0.50)		
		True*	Linear model	Tobit model	True*	Linear model	Tobit model
Treatment and time interaction	100	-0.20 (0.12)	-0.16 (0.10)	-0.20 (0.12)	-0.50 (0.12)	-0.37 (0.10)	-0.50 (0.12)
	200	-0.20 (0.08)	-0.16 (0.07)	-0.20 (0.09)	-0.50 (0.08)	-0.37 (0.07)	-0.50 (0.09)
	300	-0.20 (0.07)	-0.17 (0.06)	-0.20 (0.07)	-0.50 (0.07)	-0.37 (0.06)	-0.50 (0.07)

4 follow-up measurements

	N	MODERATE EFFECT (-0.20)			STRONG EFFECT (-0.50)		
		True*	Linear model	Tobit model	True*	Linear model	Tobit model
Treatment and time interaction	100	-0.20 (0.05)	-0.14 (0.04)	-0.20 (0.06)	-0.50 (0.05)	-0.27 (0.04)	-0.50 (0.06)
	200	-0.20 (0.04)	-0.14 (0.03)	-0.20 (0.04)	-0.50 (0.04)	-0.26 (0.03)	-0.50 (0.04)
	300	-0.20 (0.03)	-0.14 (0.03)	-0.20 (0.03)	-0.50 (0.03)	-0.27 (0.03)	-0.50 (0.03)

* Estimated treatment and time interaction in the model without zeros.

TABLE 4. Coverage* and standardized bias of the treatment and time interaction estimated in different simulated datasets*****

TREATMENT AND TIME INTERACTION		LINEAR MIXED MODEL			TOBIT MIXED MODEL			
		N=100	N=200	N=300	N=100	N=200	N=300	
2 follow-ups	Moderate effect	Coverage	93	91	92	94	95	94
		Standardized bias	33	50	43	0	0	0
	Strong effect	Coverage	77	57	43	93	95	94
		Standardized bias	108	163	186	0	0	0
4 follow-ups	Moderate effect	Coverage	75	57	44	95	96	96
		Standardized bias	120	150	200	0	0	0
	Strong effect	Coverage	0	0	0	95	95	96
		Standardized bias	460	600	767	0	0	0

N = sample size;

* The proportion of times the 95% confidence interval include the true value (values < 90 percent are in bold italic)

** Considered to be biased if standardized bias > 40 percent (in bold italic)

*** Values are compared to the true values

model analysis and a tobit mixed model analysis in a real life example, we used data from the tREACH (Treatment in the Rotterdam Early Arthritis Cohort) trial (ISRCTN26791028) [20,21]. This trial contained HAQ-DI scores of 281 patients with recent-onset arthritis that had a high probability of progressing to persistent arthritis and compared three groups with each other: Group A (91 patients) received a combination therapy (methotrexate (MTX) + sulfasalazine (SSZ) + hydroxychloroquine (HCQ)) with intramuscular glucocorticoids (IM GCs), group B (93 patients) received a combination therapy (MTX + SSZ + HCQ) with an oral glucocorticoids (oral GCs) and group C (97 patients) received MTX + oral GCs. Measurements were conducted at baseline, after 3, 6, 9, and after 12 months and because there was no linear development over time, time was treated as a categorical variable represented by dummy variables.

Table 7 shows the percentages of zeroes at every time point for the three medication groups. It can be seen that the percentages of zeroes were comparable to the ones simulated in the datasets with a moderate treatment effect, although the differences between the groups were less strong.

Figure 2 shows the predicted HAQ-DI score over time for the three medication groups. Both models showed a similar development over time. However, the tobit mixed model analysis showed a larger difference between the medication groups at 6 months (time 3) and at 9 months (time 4) in comparison to the linear mixed model analysis. The largest difference was observed at 9 months, where the tobit mixed model showed a p-value close to the .05 significance limit (i.e. 0.055) compared to the linear mixed model (i.e. 0.133).

TABLE 5. Regression coefficients and standard errors (SE) of the overall treatment effect estimated in different simulated datasets**

		2 follow-up measurements					
		MODERATE EFFECT			STRONG EFFECT		
	N	True*	Linear model	Tobit model	True*	Linear model	Tobit model
Overall treatment	100	-0.20 (0.19)	-0.16 (0.17)	-0.19 (0.19)	-0.50 (0.19)	-0.38 (0.16)	-0.48 (0.19)
	200	-0.20 (0.14)	-0.17 (0.12)	-0.20 (0.14)	-0.50 (0.14)	-0.39 (0.11)	-0.48 (0.13)
	300	-0.20 (0.11)	-0.17 (0.10)	-0.20 (0.11)	-0.50 (0.11)	-0.39 (0.10)	-0.48 (0.11)

		4 follow-up measurements					
		MODERATE EFFECT			STRONG EFFECT		
	N	True*	Linear model	Tobit model	True*	Linear model	Tobit model
Overall treatment	100	-0.40 (0.18)	-0.31 (0.15)	-0.38 (0.18)	-1.0 (0.18)	-0.63 (0.14)	-0.89 (0.18)
	200	-0.40 (0.12)	-0.31 (0.11)	-0.38 (0.13)	-1.0 (0.13)	-0.63 (0.10)	-0.90 (0.13)
	300	-0.40 (0.11)	-0.31 (0.09)	-0.38 (0.11)	-1.0 (0.11)	-0.63 (0.08)	-0.90 (0.11)

* Estimated overall treatment effect in the model without zeros.
 ** Simulation models were set up to estimate the treatment and time interaction.

TABLE 6. Coverage* and standardized bias of the overall treatment effect estimated in different simulated datasets*****

OVERALL TREATMENT			LINEAR MIXED MODEL			TOBIT MIXED MODEL		
			N=100	N=200	N=300	N=100	N=200	N=300
2 follow-ups	Moderate effect	Coverage	93	94	93	94	95	95
		Standardized bias	21	21	27	5	0	0
	Strong effect	Coverage	88	83	77	94	95	94
		Standardized bias	63	79	100	11	14	18
4 follow-ups	Moderate effect	Coverage	89	87	81	93	94	95
		Standardized bias	50	75	82	11	17	18
	Strong effect	Coverage	26	5	0	90	89	84
		Standardized bias	206	285	336	61	77	91

N = sample size;
 * The proportion of times the 95% confidence interval include the average of the true model (values < 90 percent are in bold)
 ** Considered to be biased if standardized bias > 40 percent (in bold italic)
 *** Values are compared to the estimated values from a model without zeros

TABLE 7. The HAQ-DI percentage of zeroes over time for the three treatment groups

MONTH	GROUP A	GROUP B	GROUP C
0	10%	10%	5%
3	29%	33%	23%
6	40%	36%	25%
9	39%	38%	10%
12	27%	27%	16%

HAQ-DI = Health Assessment Questionnaire Disability Index
 Group A, combination therapy (methotrexate (MTX) + sulfasalazine (SSZ) + hydroxychloroquine (HCQ)) with GCs intramuscularly (IM GCs),
 Group B, combination therapy (MTX + SSZ + HCQ) with an oral tapering scheme (oral GCs) and Group C MTX + oral GCs

DISCUSSION

The results of the present simulation study showed that the tobit mixed model performed much better than the linear mixed model in handling floor effects (an excess of zeroes) due to censoring in all conditions. When the interaction between treatment and time was estimated, the results of the tobit mixed model were the same as the true values, while the linear mixed model analysis showed underestimated regression coefficients. When the overall treatment effect was estimated, all regression coefficients were lower than the ones obtained from the linear mixed model analysis on the dataset without zeroes, but the tobit mixed model performed much better than the linear mixed model. However, when the percentage of zeroes increased

strongly over time the tobit mixed model also did not result in accurate estimates of the overall treatment effect.

According to the standardized bias and the coverage, the linear mixed model only performed sufficiently for the situation with a moderate treatment effect and only 2 follow-up measurements. However, even in this situation the treatment effects were highly underestimated. Furthermore, the empirical example showed that relevant treatment effects could be missed when analysed with a linear mixed model where a tobit mixed model would be more appropriate. It should be noted that the differences between the two methods in the TReach data were relatively small, which was due to the relatively small intervention effects.

The objective of this simulation study was to compare a longitudinal linear mixed model analysis with a longitudinal tobit mixed model analysis for different situations with floor effects due to censoring in longitudinal RCT data. The comparison was performed because in most real life situations a linear mixed model analysis is used to analyse these kind of data, ignoring the specific non-normal distribution of the data and the excess of zeroes. Not much research has been done on this topic regarding longitudinal RCT data. Twisk and Rijmen [3] conducted an empirical study and looked at the development of rehabilitation over time in stroke patients by using the Barthel index. They concluded, by examining the model fit, that the tobit mixed model was better suited to analyse outcome variables that contained ceiling effects. Wang et al. [22] have examined the performance of a linear and a tobit growth curve model in simulated data with and without ceiling effects. They showed that the use of a longitudinal linear growth model led to incorrect model selection and bias in the estimation of the magnitude of the changes over time, and that the longitudinal tobit model performed very well. In an additional empirical data analysis they found that comparing age groups with different proportions of ceiling data could lead to wrong conclusions when traditional methods were applied. We found similar results in our simulation study regarding floor effects and showed that the use of linear mixed model analyses led to a high underestimation of the interaction between treatment and time.

When the overall treatment effect was considered, the linear mixed model analysis again showed a strong underestimation of the treatment effect. However, also the tobit mixed model analysis showed a slight underestimation of the treatment effect, which became stronger when the percentage of zeros increased. It should be realised that the overall treatment effect was estimated from the same simulation models as the interaction between treatment and time and that the simulation models were set up to estimate the interaction between treatment and time. In an additional simulation study, we set up the simulation models to estimate the overall treatment effect, without adding an interaction between treatment and time. In these simulations, the regression coefficients for the overall

treatment effect of the tobit mixed model analyses were the same as the true values, while the regression coefficients of the linear mixed model analyses were (again) much lower (see Table 8).

It should be realized that the tobit mixed model analysis is one of the options to deal with the excess of zeros in RCTs. There are other possibilities to deal with this problem. Specifically for RCT's an adjustment for the baseline value of the outcome often leads to a normal distribution of the residuals even in situations where the outcome variable is hampered by an excess of zeros. Adjustment for the baseline value of the outcome is often used to take into account differences in baseline values between the treatment and control group [2,23]. However, when the outcome variable at baseline is not hampered by an excess of zeros, while the follow-up measurements are, it is doubtful whether the residuals of the analyses will be normally distributed. Besides the tobit model, other regression models also can be considered to deal with the excess of zeros [24]. An extensive discussion of these (complicated) methods is beyond the scope of this paper.

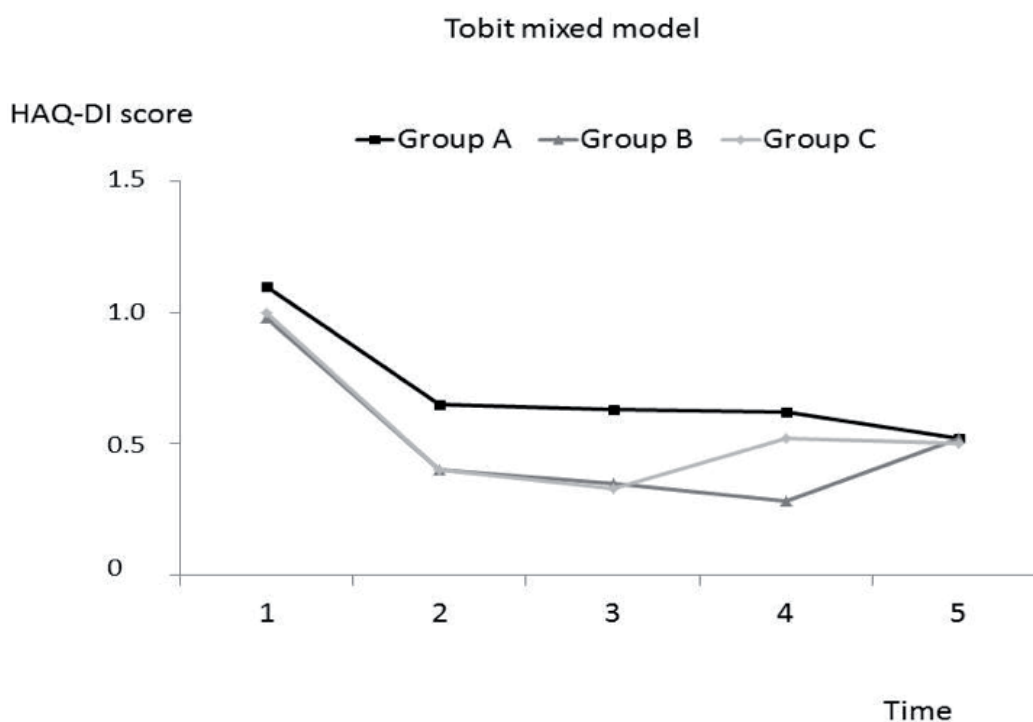
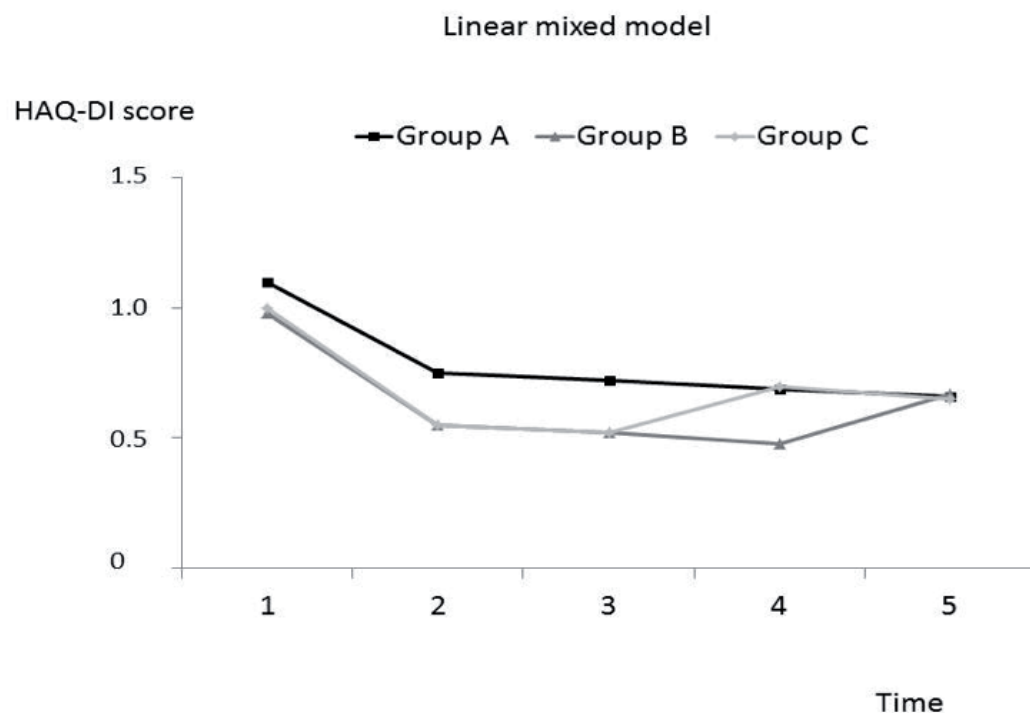
In the present simulation study, we analysed two possible treatment effects. The moderate treatment effect of 0.2 would be the most realistic to occur in real life data. We included the stronger treatment effect to investigate the model performance in more extreme situations. For both the moderate and the strong treatment effect, the use of a tobit mixed model analysis led to more valid results than a linear mixed model analysis. Moreover, we simulated a 'true' outcome without floor effects that was normally distributed, after which floor effects were created by setting all negative values to zero. Thus, we assumed that the true outcome is normally distributed. It should be noted that a stronger treatment effect that remains for a longer period over time influences the shape of the distribution. It is important to keep in mind that real life outcomes with floor effects may not always follow an underlying normal distribution. It is not clear how the tobit mixed model will behave in those situations.

In this paper we assumed that the excess of zeroes in the outcome was due to censoring. However, an excess of zeroes does not always occur because of censoring, zeroes in the data may reflect 'real' zeroes. Examples of such data are the number of cigarettes smoked to analyse smoking behaviour, or the number of hypoglycaemic events in diabetic patients. The tobit model is not suitable to analyse these kinds of data. Outcomes that contain 'real' zeroes can better be analysed with other methods [25,26].

CONCLUSION

The tobit mixed model performed much better than a linear mixed model and should therefore be used in longitudinal RCTs with floor effects due to censoring.

FIGURE 2. Predicted HAQ-DI score over time for the three treatment groups



HAQ-DI: Health Assessment Questionnaire Disability Index;
 Group A: MTX (methotrexate) + SSZ (sulfasalazine) + HCQ (hydroxychloroquine (HCQ)) with IM GCs (intramuscular glucocorticoids)
 Group B: MTX + SSZ + HCQ + with oral GCs (oral glucocorticoids)
 Group C: MTX + oral GCs.

TABLE 8. Regression coefficients and standard errors (SE) of the overall treatment effect (true effect equals -0.5) estimated for simulated datasets.

2 FOLLOW-UP MEASUREMENTS				4 FOLLOW-UP MEASUREMENTS			
N	True*	Linear model	Tobit model	N	True*	Linear model	Tobit model
100	-0.50 (0.19)	-0.40 (0.16)	-0.50 (0.19)	100	-0.50 (0.18)	-0.38 (0.14)	-0.50 (0.18)
200	-0.50 (0.14)	-0.40 (0.11)	-0.50 (0.14)	200	-0.50 (0.13)	-0.38 (0.10)	-0.50 (0.13)
300	-0.50 (0.11)	-0.41 (0.09)	-0.50 (0.11)	300	-0.50 (0.11)	-0.38 (0.08)	-0.50 (0.11)

* Estimated overall treatment effect in the model without zeros.

Funding information

The iREACH study was funded by an unrestricted grant from Pfizer bv. (0881–102217).

References

- Bothwell LE, Greene JA, Scott H, Podolsky SH and Jones DS. Assessing the Gold Standard — Lessons from the History of RCTs. *N Engl J Med* 2016; 374: 2175-2181.
- Twisk J, Bosman L, Hoekstra T, Rijnhard J, Welten M, Heymans M. Different ways to estimate treatment effects in randomised controlled trials. *Contemporary Clinical Trial Communications*, 2018, 10, 80-85.
- Twisk J, Rijmen F. Longitudinal tobit regression: A new approach to analyze outcome variables with floor or ceiling effects. *J Clin Epidemiol*. 2009;62:953-8.
- Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. *Health Qual Life Outcomes*. 2003;1:20.
- McElduff F, Cortina-Borja M, Chan S-K and Angie Wade A. When t-tests or Wilcoxon-Mann-Whitney tests won't do. *Advances in Physiological Education* 2010; 34: 128–133.
- Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. *Biometrics*. 1982;38:963-74.
- Twisk JWR. *Applied longitudinal data analysis for epidemiology. A practical guide*. Second edition. Cambridge: Cambridge University Press; 2013.
- Ledbetter MF, Smith LA, Vosler-Hunter WL, Fischer JD. An evaluation of the research and clinical usefulness of the Spiritual Well-Being Scale. *Journal of Psychology and Theology*. 1991; 19: 49-55.
- Murrell RC, Kenealy PM, Beaumont JG, Lintern TC. Assessing quality of life in persons with severe neurological disability associated with multiple sclerosis: The psychometric evaluation of two quality of life measures. *British Journal of Health Psychology*. 1999;4:349-62.
- Genia V. Evaluation of the Spiritual Well-Being Scale in a Sample of College Students. *The International Journal for the Psychology of Religion*. 2001;11:25-33.
- McBee M. Modeling Outcomes With Floor or Ceiling Effects: An Introduction to the Tobit Model. *Gifted Child Quarterly*. 2010;54:314-20.
- Tobin J. Estimation of Relationships for Limited Dependent-Variables. *Econometrica*. 1958;26:24-36.
- Hajivassiliou VA. A simulation estimation analysis of the external debt crises of developing countries. *Journal of applied econometrics*. 1994;9:109-31.
- Kyriazidou E. Estimation of a panel data sample selection model. *Econometrica: Journal of the Econometric Society*. 1997:1335-64.
- Stata Corporation. *Stata Statistical Software*. 14.0 ed. College Station: Stata Press; 2016.
- Rabe-Hesketh S, Skrondal A. *Multilevel and longitudinal modeling using stata*. College Station: Stata Press; 2005.
- Kang JH. The usefulness and uselessness of the decomposition of tobit coefficients. *Social Method Res*. 2007;35:572-82.
- Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med*. 2006;25:4279-92.
- Collins LM, Schafer JL, Kam C-M. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*. 2001;6:330.
- Claessen SJ, Hazes JM, Huisman MA, van Zeven D, Luime JJ, Weel AE. Use of risk stratification to target therapies in patients with recent onset arthritis; design of a prospective randomized multicenter controlled trial. *BMC Musculoskelet Disord*. 2009;10:71.
- de Jong PH, Hazes JM, Barendregt PJ, Huisman M, van Zeven D, van der Lubbe PA, et al. Induction therapy with a combination of DMARDs is better than methotrexate monotherapy: first results of the iREACH trial. *Ann Rheum Dis*. 2013;72:72-8.
- Wang U, Zhang ZY, McArdle JJ, Salthouse TA. Investigating ceiling effects in longitudinal data analysis. *Multivar Behav Res*. 2008;43:476-96.
- Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *BMJ* 2001;323:11234.
- Stasinopoulos MD, Rigby RA, Heller GZ, Voudouris V, De Bastiani F. *Flexible Regression and Smoothing: Using GAMLSS in R*. CRC Press, 2017
- Liu H, Powers DA. Growth curve models for zero-inflated count data: An application to smoking behavior. *Struct Equ Modeling*. 2007;14:247-79.
- Spriensma AS, Hajos TRS, de Boer MR, Heymans MW, Twisk JWR. A new approach to analyse longitudinal epidemiological data with an excess of zeros. *BMC Med Res Methodol*. 2013;13.

