

Analyzing Proportion Scores as Outcomes for Prevention Trials: a Statistical Primer

Kehui Chen^{1,2} · Yu Cheng^{1,2} · Olga Berkout² · Oliver Lindhiem²

Published online: 10 March 2016
© Society for Prevention Research 2016

Abstract In prevention trials, outcomes of interest frequently include data that are best quantified as proportion scores. In some cases, however, proportion scores may violate the statistical assumptions underlying common analytic methods. In this paper, we provide guidelines for analyzing frequency and proportion data as primary outcomes. We describe standard methods including generalized linear regression models to compare mean proportion scores and examine tools for testing normality and other assumptions for each model. Recommendations are made for instances when the assumptions are not met, including transformations for proportion scores that are non-normal. We also discuss more sophisticated analytical tools to model change in proportion scores over time. The guidelines provide ready-to-use analytical strategies for frequency and proportion data that are commonly encountered in prevention science.

Keywords Proportion scores · Data transformations · Generalized linear models · Beta regression · Prevention trials

In prevention studies, we often encounter proportion data. For example, instruments are often divided into several domains or subscales, and proportions of subscale scores are computed relative to the total scores (e.g., Chamberlain et al. 2008; Lindhiem et al. 2014; Schuhmann et al. 1998). Due to the intrinsic nature of proportions, it is critical to analyze such data

properly. In this paper, we review existing methods for analyzing proportional data and provide practical statistical guidelines. Moreover, we illustrate how to apply linear and generalized mixed-effects models to longitudinal proportion data to model changes in proportion scores or frequencies over time. In order to enhance the integration of this special series, these concepts and analytic methods are illustrated using proportion data from the Services for Kids in Primary Care (SKIP2) study described in this special series (Shaffer et al. [this issue](#)).

Overview of Proportion Data

The main feature of proportion data is that they are bounded by the interval $[0, 1]$. Consequently, the data often exhibit heterogeneity in variance (i.e., the variability tends to be higher for data values in the middle range than that toward the boundaries). When studying the association with a continuous predictor, the relationship often deviates from linearity toward the boundaries. If one ignores this special feature of proportion data, the untransformed proportion data can be analyzed using standard methods such as the analysis of variance (ANOVA) and ordinary least squares regression. However, some statistical assumptions underlying these models, such as linearity, normality, and homogeneous errors, might be violated due to these special features of proportion data. Another issue with untransformed data is that the predicted values, as well as prediction intervals, could give rise to values outside of the range $[0, 1]$. A common practice to resolve these issues is to transform proportion data. While transformation can map $[0, 1]$ data onto unbounded real values, this practice does not take the intrinsic probability characteristics of proportion data into account. Furthermore, studies have

✉ Oliver Lindhiem
lindhiemoj@upmc.edu

¹ Department of Statistics, University of Pittsburgh, Pittsburgh, PA, USA

² Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

shown that data transformations might be inefficient with lower power, and, perhaps more importantly, the results are difficult to interpret (e.g., Warton and Hui 2011).

The generalized linear model, an alternative to the transformation methods, has become more prevalent in analyzing proportion data in clinical studies. The model naturally handles non-normality and heterogeneity issues, and the use of a link function guarantees that the fitted values will be exactly within the desired range $[0, 1]$. For example, logistic regression, a common generalized linear model, is suitable for a binomial outcome, where the proportion is computed as the ratio of the number of target events to the total number of trials, “ n_y out of n .” The fitted logistic model is often more interpretable than a transformed model and reflects the underlying characteristics of the data better than linear regression. For another type of proportional data, where the values are between 0 and 1 but are not exactly binomial, one may use an over-dispersed logistic model or beta regression model, which allows more flexible variance structures.

To illustrate the methods, throughout the paper, we will use sample proportion data from the SKIP2 study (Kolko et al. 2014; Shaffer et al. [this issue](#)). One of the goals of the SKIP2 program was to examine how a Doctor-Office Collaborative Care (DOCC) treatment improves parent-reported disciplinary practices as compared to Enhanced Usual Care (EUC). For illustration purpose, in this paper, we focus primarily on the proportion of using corporal punishment among all disciplinary practices, estimated from the Alabama Parenting Questionnaire (APQ; Shelton et al. 1996), as this proportion is in general smaller than 0.2 and serves to illustrate many of the concepts and guidelines. We will first examine the analysis of these proportion data using linear regression, the transformation methods, and the generalized linear models. Then, we will briefly discuss linear and generalized linear mixed-effects models for evaluating the changes in these proportion data over both the acute and maintenance phases of treatment.

Regression Analysis

Linear Models for Untransformed Data

Let us consider a common case, where for each subject (i), we observe some outcome or response (Y_i) ranging between 0 and 1 and explanatory variables (X_i). In this case, the linear model for our untransformed data would be

$$Y_i = X_i^T \beta + \epsilon_i,$$

where X_i represents values for our explanatory variables, β the coefficients describing the linear relationship between X_i and Y_i , and ϵ_i the error in predicting our outcome of interest. Error values are assumed to be independent and identically

distributed, that is, values for particular individuals are not expected to be related to those of others (e.g., they are not influenced by a common cause such as systematic measurement error) and they are expected to share a common distribution. ANOVA is a special case of this kind of model, where the explanatory variables are categorical. The main concerns of modeling untransformed data are twofold. First, proportion data often violate model assumptions, including linearity, normality, and homogeneity of errors, especially for data points that are toward the boundaries. The distribution of Y is usually a mixture of distributions with different means for all the observations. Therefore, we would like to emphasize that the check of normality is for the residual values. Second, the model does not take into account the range restriction inherent in use of proportion scores and predicted values and intervals could fall outside of $[0, 1]$. In practice, one needs to first examine the histogram of the response data Y and check the residual plots after model fitting. Since the deviation from linearity and constant variance for proportion data is most severe toward boundaries 0 and 1, it is relatively safe to use untransformed normal model for datasets that values fall mainly between 0.2 and 0.8, although the residual distribution should also be examined. Plots describing the residual distribution can be obtained in most statistical packages, such as SAS, R, Stata, and SPSS. The Q-Q plot is a quantile plot, which compares the sorted obtained residual values with quantiles from a normal distribution. If we observe an approximate straight line using the Q-Q plot, we obtain support for the assumption that our residuals are normally distributed.

In addition to the Q-Q plot, we can also look at a residual plot compared to the predicted plot. This residual vs. predicted plot compares the value of our residuals to predicted values of our outcome variable and shows us whether there is greater variability in error at higher or lower values. Generally, as our model assumes that errors are identically distributed, we hope that the spread of error appears similar across predicted outcome values. Statistical tests of normality, such as Shapiro-Wilk (available in SAS, R, Stata, and SPSS), may also be performed to assess whether the distribution of error values deviates from normality. However, findings of statistical significance are more likely in larger samples, and visual examination of plots should also be conducted. If the assumptions described above are met, it is usually safe to proceed with untransformed data, but appropriate diagnostics (see examples below) are still necessary as in regular linear regression. If the residual distribution is substantially skewed, however, our statistical tests and confidence intervals (based on the normal assumption) will be invalid.

In our data example, there are three outcome values, PROPcp (proportions of corporal punishment), PROPPp

(proportion of positive parenting), and PROPid (proportion of inconsistent discipline). By definition, the three proportions add up to 1. The outcomes were measured longitudinally at baseline and again at 6, 12, and 18 months after baseline. We first would like to evaluate the short-term treatment effect by focusing on the proportions at time 2 (i.e., after 6 months of treatment). Figure 1 shows the histograms of the three outcome variables (from left to right: PROPCp2, PROPPp2, and PROPid2). We can see that the distributions of PROPPp2 and PROPid2 are quite normal and their values range mainly from 0.2 to 0.8. As we mentioned earlier, it is relatively safe to proceed with untransformed normal models for PROPPp2 and PROPid2. We hence fit linear regression models for these two outcomes on the primary variable of interest, the group assignment, while accounting for age and gender, as well as the corresponding proportion score at baseline. The residual plots for these two outcome variables look quite reasonable (see Fig. 2).

In contrast, PROPCp2 is obviously right skewed with most values between 0.05 and 0.2 (compared to PROPPp2 and PROPid2 in Fig. 1). Nevertheless, we performed linear regression for the untransformed data for illustration purpose (Table 1). One data point with PROPCp2 of 0.25 was identified as an outlier from regression diagnostic plots and was removed from the current analysis (further detail on data cleaning and examination for regression analyses can be found in Tabachnick and Fidell 2007). Information on residuals and predicted and observed values of PROPCp2 can be seen in Fig. 3. The residual vs. predicted value plot in the left panel of Fig. 3 suggests that as PROPCp2 values increase, so does their variance

(termed a variance on mean relationship). The normal Q-Q plot for residuals (the middle panel of Fig. 3) shows a slight deviation from normality, although the Shapiro-Wilk test for normality of residuals is not significant.

On the right panel of Fig. 3, we plot the observed and predicted PROPCp2 against the linear predictor $X\hat{\beta}$ (representing the effect of treatment). Large values among the observed and predicted PROPCp2 scores do not match well. After accounting for significant prediction by PROPCp1 and other variables, the influence of DOCC treatment on PROPCp2 is not significant compared to EUC.

Transformation Model

To partially overcome some of the data issues mentioned above, one possibility is to transform the data. The transformation model assumes that

$$h(Y_i) = X_i^T \beta + \epsilon_i,$$

where h is some known transformation function (e.g., such as taking the square root of Y). In some fields, such as ecology, the arcsine square root transformation has long been a common practice to transform proportional data. Arcsine square root can stabilize the variance when the proportional data are of binomial distribution, computed as “ n_y out of n ” (Gotelli and Ellison 2004; Sokal and Rohlf 1995; Zar 1998). However, the arcsine square root transformation is not monotone, which makes the interpretation of coefficients obscure. Monotone functions preserve the rank order of data points, and the absence of such ordering can make it difficult to draw

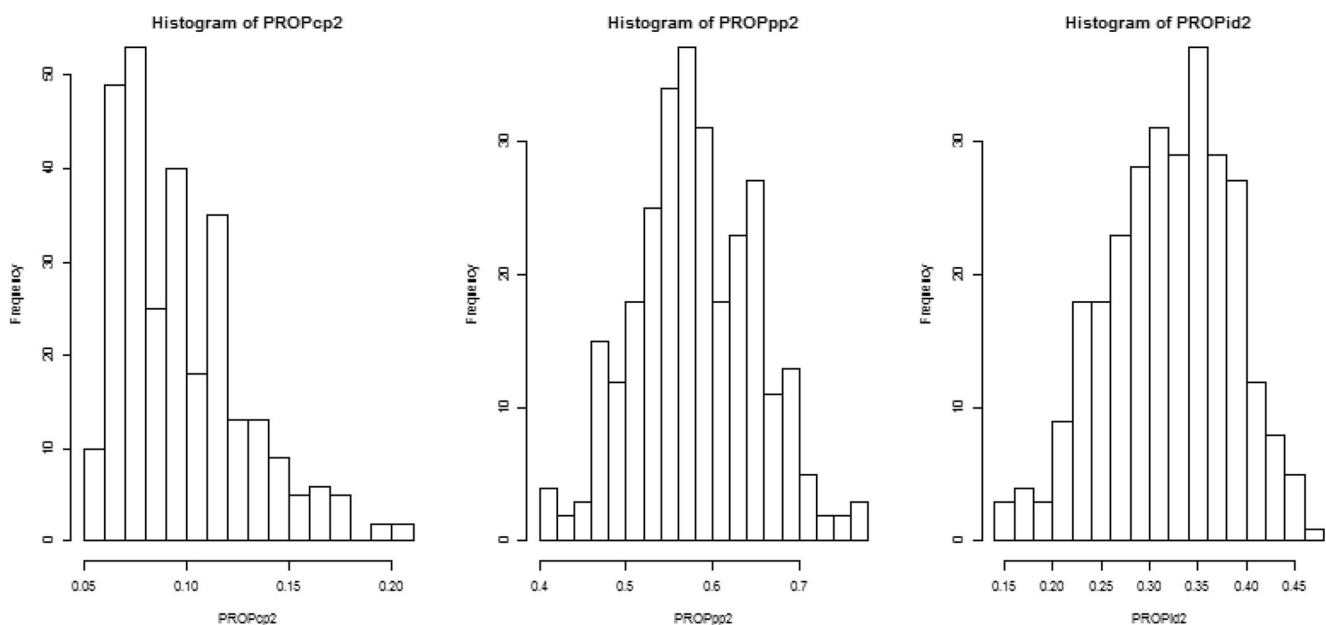
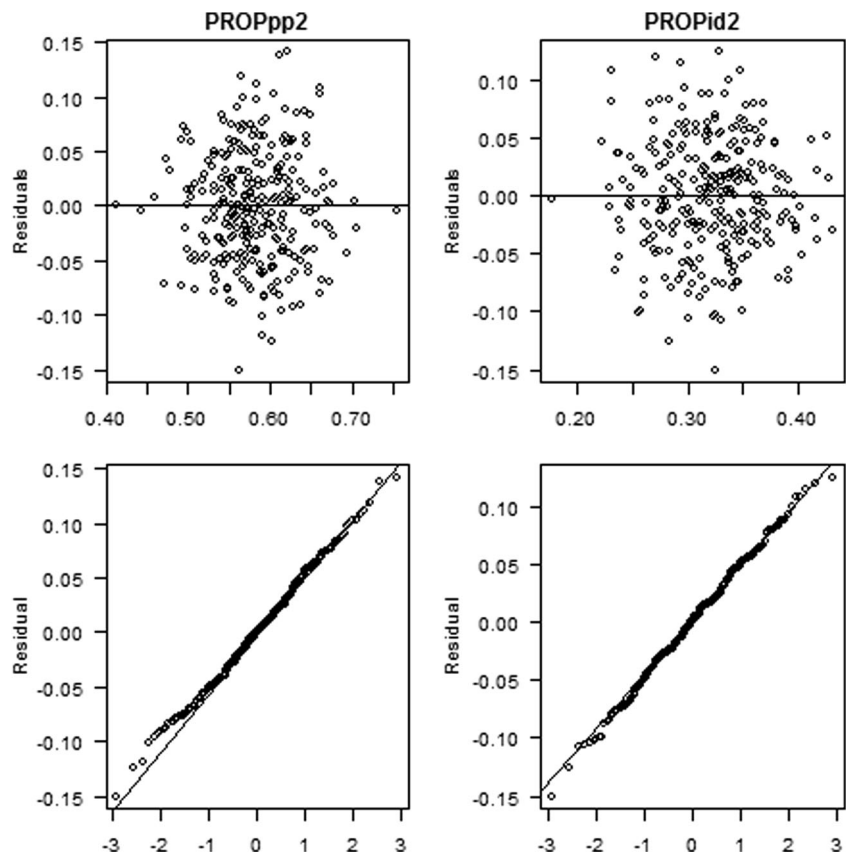


Fig. 1 Histograms for three outcomes. *PROPCp2* corporal punishment proportion score at 6-month assessment, *PROPPp2* positive parenting proportion score at 6-month assessment, *PROPid2* inconsistent discipline proportion score at 6-month assessment

Fig. 2 Residual plots for PROPpp2 (left) and PROPid2 (right). PROPpp2 positive parenting proportion score at 6-month assessment, PROPid2 inconsistent discipline proportion score at 6-month assessment



conclusions from transformed data. Also, several recent papers (Warton and Hui 2011; Wilson and Hardy 2002) have reported that logistic regression (discussed in the next section) is more powerful than arcsine square root for binomial proportional data. Another popular transformation is the complementary log-log, i.e., $\log(-\log(1-p))$, which is a monotone function widely used for transforming percentile data in the survival analysis literature.

For illustration, we applied the arcsine square root and complementary log-log transformations to the PROPcp2 data. The residual plots and normal Q-Q plots for regression after the arcsine square root transformation are shown in the top left panel of Fig. 4. The residual vs.

predicted value plot has improved slightly over the untransformed model (as reflected by less residual increase at higher predicted values, Fig. 4 and Table 2) but still shows some heteroscedasticity in variance. This is as expected, since our proportion data are not binomial. The complementary log-log transformation is found to be a little bit better than the arcsine square root transformation in terms of stabilizing variance, as seen in the residual plots and Q-Q normal plots for linear regression after the complementary log-log transformation (Fig. 4, right panel).

In practice, it is often difficult to find a reasonably simple transformation that could simultaneously stabilize

Table 1 Estimation for untransformed normal model (left) and beta regression with identity link (right)

Variable	Untransformed normal model					Beta regression with identity link				
	Estimate	Standard error	df	t-value	Prob> t	Estimate	Standard error	df	t-value	Prob> t
Intercept	0.020	0.008	1	2.57	0.011	0.022	0.008	251	2.93	0.004
Child age	0.000	0.001	1	0.71	0.481	0.001	0.001	251	0.92	0.359
Child gender	-0.001	0.003	1	-0.54	0.589	0.000	0.003	251	0.10	0.922
PROPcp1	0.719	0.043	1	16.73	<0.000	0.674	0.043	251	15.61	<0.000
Treatment condition	-0.004	0.003	1	-1.50	0.134	-0.002	0.003	251	-0.87	0.385

df degrees of freedom, Prob probability, PROPcp1 corporal punishment proportion score at baseline

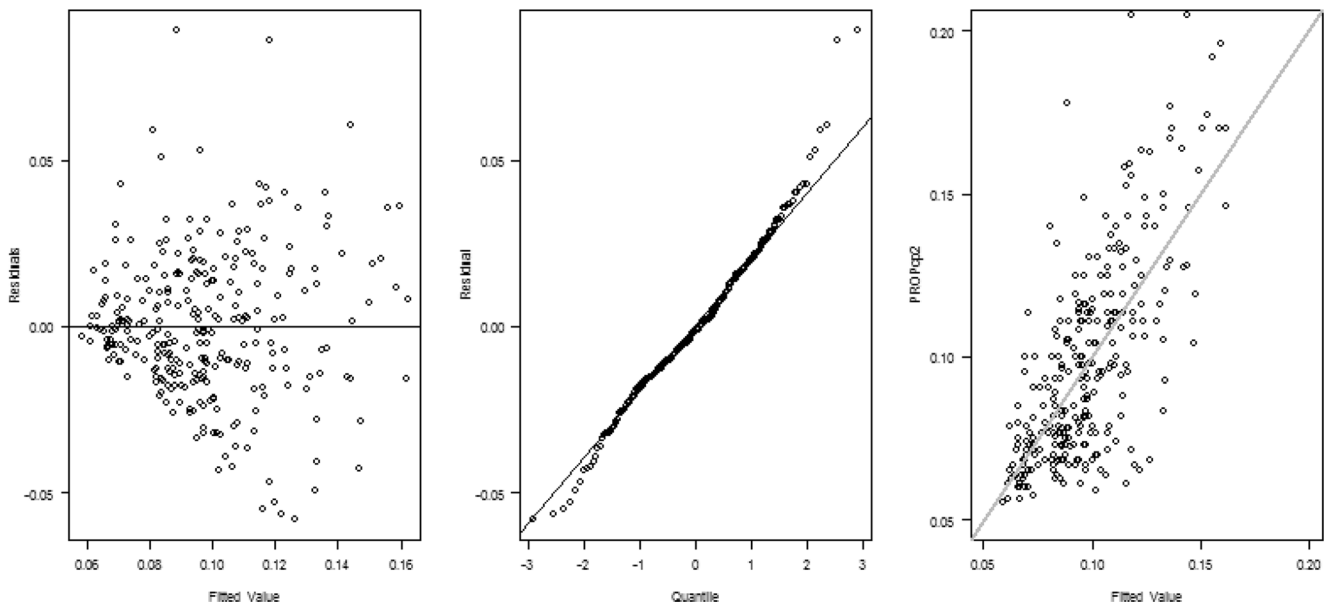


Fig. 3 Untransformed normal model for corporal punishment proportion score at 6-month assessment (PROPcp2)

the error variance, ease the violation of normality for the error term, and maintain the modeling efficiency. Perhaps more importantly, coefficients may only be interpreted using the transformed values, a major consideration for drawing conclusions. Software outputs, such as predicted values, confidence intervals, and prediction intervals, are

all in the transformed scale and need to be transformed back to the original scale. The estimated coefficients are shown in Table 2. The *R*-squared values for both transformed models are still 0.5. The significance results remain the same, while the fitted coefficients are not directly comparable to the untransformed normal model.

Fig. 4 Results for transformation methods: arcsinesq (left) and complementary log-log (right)

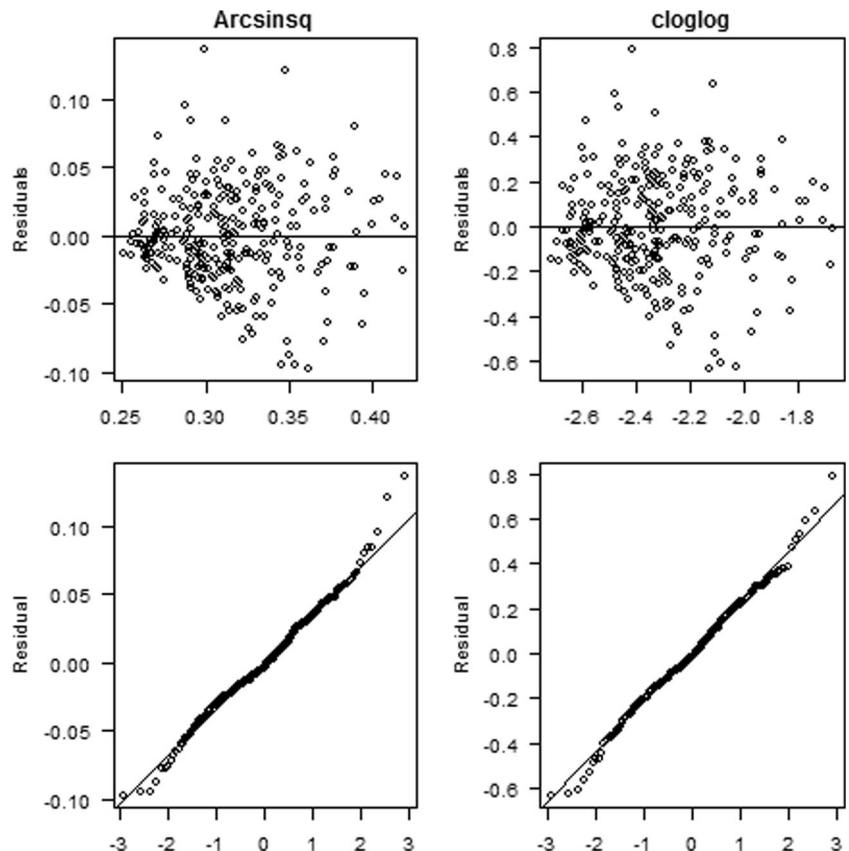


Table 2 Results for transformation methods: arcsineq (left) and complementary log-log (right)

Variable	Arcsineq					c-log-log				
	df	Parameter estimate	Standard error	t-value	Prob > t	df	Parameter estimate	Standard error	t-value	Prob > t
Intercept	1	0.187	0.013	14.76	<0.000	1	-3.124	0.080	-39.14	<0.000
Child age	1	0.001	0.001	0.84	0.399	1	0.007	0.007	0.99	0.325
Child gender	1	-0.002	0.004	-0.52	0.601	1	-0.014	0.028	-0.51	0.609
PROPcp1	1	1.167	0.070	16.75	<0.000	1	7.256	0.439	16.54	<0.000
Treatment condition	1	-0.006	0.004	-1.35	0.177	1	-0.031	0.027	-1.16	0.247

df degrees of freedom, Prob probability, PROPcp1 corporal punishment proportion score at baseline

Generalized Linear Models

Generalized linear models (Dobson 2002; McCullagh and Nelder 1989) are a family of models that allow the response variables to have non-normal distributions, such as exponential, binomial, Poisson, gamma, and beta distributions. These can be thought of as a broader and more general application of linear regression, where a link function (an equation) is added to the general linear model to describe the relationship between the predictors and their coefficients and outcome (Olsson 2002).

The normal linear regression is a special case of generalized linear models with the identity link function and normal distribution. The identity link is a special kind of link function, used when data are suitable for the general linear model analysis, whereas in generalized linear models, others may be used depending upon distribution. Logistic regression is a popular type of generalized linear model, where the outcome distribution is binomial and the link function is the logit. The logistic regression model is arguably the best model for proportion data when it is computed as “ n_y out of n ” with $Y = \frac{n_y}{n}$. The number of events is then modeled as binomial, where the explanatory variables contribute to its prediction through use of a logit link function. The coefficients in logistic regression have a natural interpretation, where one unit increase in the predictor X leads to an increase in the odds (probability an event will occur/1 probability event will occur) by a factor of e^β , where β is the regression coefficient for X . When the data under consideration are proportions of diagnoses or events out of the total of instrument items or trials, logistic regression correctly models binomial data and usually is more efficient and powerful than transformation methods (Warton and Hui 2011). The use of a logit link function guarantees that the fitted values will be exactly within the desired range [0, 1].

We would like to note that the use of a logit link function is different from linear regression after logit transformation. The logit link in logistic regression is an integrated modeling component, which allows the mean function to be non-linear. No transformation is performed on the actual data Y , and there is no requirement for normality in any forms. However, in the logit

transformed linear regression, we still require approximate normality after transformation, and moreover, the logit transformation of data points might encounter problems when there are exact zeros or ones in the Y and adjustments are needed for these cases.

When the data are not exactly binomial and there is variability in outcome beyond that predicted by the model, data over-dispersion presents problems (Raim and Neerchal 2013). More specifically, over-dispersion for proportion data $Y = \frac{n_y}{n}$ means that the variance is greater than $\frac{\pi(1-\pi)}{n}$ as would be expected based on a binomial model, where we assume $\pi = E(Y)$ has been correctly modeled. Over-dispersed binomial models account for this issue by including an additional parameter to more accurately describe outcome variance. On the other hand, some proportion data are continuous, such as our proportion score for corporal punishment, and are better modeled as a continuous variable rather than modeling n_y as discrete binomial variable. This type of outcome may best be handled using a beta regression. Beta regression is a generalized linear model subtype which can handle outcomes ranging from 0 to 1 and uses a beta distribution to model outcome (Ferrari and Cribari-Neto 2004). If Y also assumes the extremes 0 and 1, a useful adjustment in practice is $(Y \times (n - 1) + 0.5) / n$, where n is the sample size (Cribari-Neto and Zeileis 2010; Smithson and Verkuilen 2006). The flexibility inherent in the beta distribution makes this analytic technique particularly suitable for proportion data; this is because the shape of the beta distribution can vary widely depending upon two parameters (Ferrari and Cribari-Neto 2004). The beta distribution can be symmetric, asymmetric, including left- and right-skewed, flat-, or bell-shaped (similar to normal but restricted to values between 0 and 1). The flexibility makes the beta distribution an attractive alternative for modeling proportion data when the exact distribution is unknown.

When a logit link function is used, beta regression has the same mean and variance relationship as an over-dispersed logistic regression, and the coefficients of covariates follow a similar interpretation as in logistic regression. Moreover, fitting beta regression does not

depend on counting n_y out of n , making it suitable for more general proportional and percentile data. We fit the beta regression model with the logit link for our data example. In Fig. 5, top panel, we plot the Pearson residual (Cribari-Neto and Zeileis 2010) vs. predicted values. This is a residual plot that mimics the residual plot in linear regression and can be used for model diagnostics. In the bottom panel of Fig. 5, the predicted PROPcp2 is plotted as a function of the linear coefficients and predictor variables, overlaid with the observed PROPcp2. Pearson residual vs. predicted values in Fig. 5 can be compared with the PROPcp2 vs. predicted values shown in the right panel of Fig. 3 of the untransformed model. In the untransformed model, the linear predictor is the same as the predicted values of Y , since the link function is an identity function. We can see that both the residual plot and the

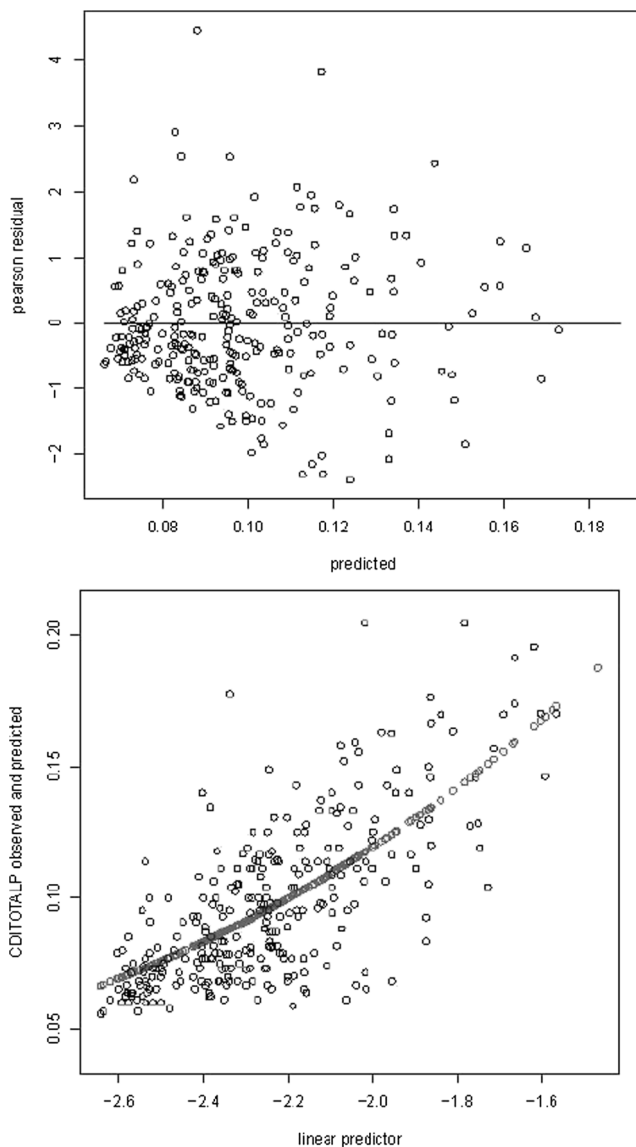


Fig. 5 Beta regression with logit link function

fitted value plot are improved compared to the untransformed normal model. The fitted coefficients are shown in Table 3. The fitted coefficients are not directly comparable to the untransformed normal model due to use of the logit link function. Let $\mu = E(Y)$ be the expected proportion of corporal punishment. The model with logit link establishes that $\log\frac{\mu_i}{1-\mu_i} = \mathbf{X}_i^T \boldsymbol{\beta}$. Following the convention in the logistic regression, $\frac{\mu_i}{1-\mu_i}$ is the odds of corporal punishment, and e^{β_j} represents the ratio of the odds of corporal punishment when the j th covariate takes on a value of $x_{ij} + 1$ to the odds when the covariate is x_{ij} , while fixing other covariates. Table 3 shows that the estimated coefficient for the only significant predictor PROPcp1 is 1.167, which means that the odds of corporal punishment at 6 months is expected to increase by a factor $e^{1.167} = 3.2$ for one unit increase in the baseline value of PROPcp. Usually, it makes more sense to use a link function that maps the linear predictor onto the desired range $[0, 1]$. However, for illustration purpose, we also fit beta regression with identity link and the fitted coefficients are shown in the right panel of Table 1. The estimated coefficients are similar to the untransformed normal model. Only the baseline PROPcp score is found to be significant. The generalized linear models can be fitted using *proc glimmix* in SAS by specifying appropriate link functions and distributions such as binomial or beta. R function *glm* (“stat”) can be used for logistic regression, and R function *betareg* (“betareg”) can be used for beta regression.

Analysis for Longitudinal Proportional Data

Linear Mixed Models

We now focus on the longitudinal feature of the proportion data. In our data example, the PROPcp score was administrated at baseline and again at 6, 12, and 18 months after baseline. The primary goal of this longitudinal analysis is to evaluate treatment efficacy over

Table 3 Coefficients for beta regression with logit link function

Effect	Estimate	Standard error	df	t-value	Prob> t
Intercept	-3.069	0.084	280	-36.57	<0.000
Child age	0.006	0.007	280	0.91	0.361
Child gender	-0.013	0.029	280	-0.46	0.648
PROPcp1	7.503	0.439	280	17.11	<0.000
Treatment condition	-0.038	0.028	280	-1.37	0.170

df degrees of freedom, Prob probability, PROPcp1 corporal punishment proportion score at baseline

time. As one of the outcomes of interest, we would like to examine if parents in DOCC clinics reported more reduction in the relative frequency of using corporal punishment than those parents from EUC clinics (see Shaffer et al. [this issue](#)).

Linear mixed-effects models are often the primary analytical tools for analyzing repeated measures with normally distributed errors. The use of linear mixed models for proportion data is generally inappropriate for essentially the same reason we have discussed in previous sections. We start with the linear mixed model here mainly for comparative and pedagogical purposes. The more suitable generalized linear mixed-effects models will be discussed in the next section. We briefly describe the linear mixed-effects model in the following, and readers who are interested in further details are referred to Jiang (2007), Littell et al. (2006), and Pinheiro and Bates (2000). As suggested by the name, mixed-effects models include both “fixed” and “random” effects. The “fixed” terms are similar to explanatory variable terms in linear and generalized linear regression. “Random effects” terms are added to the model to handle the dependence among repeated measures and to model between- and within-subject response variability. Another way to think about these is that fixed effects are the same across subjects, but random effects can differ between individuals (Liu et al. 2012). Linear mixed-effects models are suitable for longitudinal data analysis, as they allow for both baseline and time varying variable influence to be modeled.

The mixed-effects models can also deal with missing data, by including all subjects in the analysis as long as they have at least one data point, under the assumption of missing at random (MAR). MAR requires that missingness of data be accountable for by information within the dataset but allows for missingness of data to depend upon available variables. Per recommendations by Schafer and Graham (2002) and Graham (2009), the MAR assumption is reasonable for this dataset. Therefore, we first fit the mixed-effects models using all available data points, assuming normal distributions for terms representing error. Fixed terms include group assignment (SettingFam1), time as a factor, and the time by group interaction, as well as baseline predictors

including child age and child gender. Subject is included as a random term to account for the dependence among repeated measures from the same subject.

The type I ANOVA table from the above linear mixed model is given in Table 4. As in a typical longitudinal study, there are missing values in our data example. We chose to compute the denominator degrees of freedom using the Kenward-Roger method, since Alnosaier (2007) and Gregory (2011) demonstrated the superiority of the Kenward-Roger method over other popular methods, such as the Satterthwaite approximation and the containment method (interested readers can obtain further information in Alnosaier (2007) and Gregory (2011)). After adjusting for child age and child gender, there is a highly significant time effect, $F(3, 843) = 23.7, p < 0.0001$. However, neither the group nor the group by time interaction is significant. This suggests that parents in both the DOCC and EUC clinics improved over time almost equally in the use of corporal punishment (see Shaffer et al. [this issue](#)). The change from visit 1 to visit 2 (during the acute treatment period) is significant, $t(849) = -6.57, p < 0.0001$. The change from 6 to 18 months after baseline is not statistically significant, $t(836) = -0.31, p = 0.76$. Linear mixed models can be fitted by *proc mixed* or *proc glimmix* (with normal distribution and identity link) in SAS and function *lmer* (“lme4”) in R.

Generalized Linear Mixed Models

Generalized linear mixed models (or GLMMs) are an extension of linear mixed models to allow response variables from different distributions, such as binomial and beta distributions, which are useful for proportion data modeling. Alternatively, one could think of GLMMs as an extension of generalized linear models to include both fixed and random effects. The discussions about linear mixed models are all applicable to generalized linear mixed models. If the proportional data are computed from binomial variable, we can use mixed-effects logistic regression, which allows for random effects to be included in the logistic regression model and to account for correlations within subject data and over repeated measures. If the proportion data

Table 4 Type 1 tests for linear mixed model

Effect	Num df	Den df	F-value	Prob > F
Child age	1	315	1.35	0.247
Child gender	1	315	1.17	0.281
Treatment condition	1	315	1.05	0.307
Time	3	843	23.73	<0.000
Treatment condition × time	3	843	1.19	0.313

df degrees of freedom, *Prob* probability, *Num* numerator, *Den* denominator

Table 5 Type 1 tests for mixed-effects beta model with logit link function

Effect	Num df	Den df	F-value	Prob > F
Child age	1	321	1.22	0.269
Child gender	1	320	1.04	0.309
Treatment condition	1	321	1.00	0.317
Time	3	849	25.24	<0.000
Treatment condition × time	3	849	1.36	0.253

df degrees of freedom, *Prob* probability, *Num* numerator, *Den* denominator

are not binomial, as discussed above, we could consider overdispersed binomial, as well as the beta distribution.

We illustrate the mixed-effects beta regression with the logit link function using the proportions of corporal punishment as the outcome. The fitted coefficients should be interpreted in view of the logit link function and are not directly comparable to the untransformed linear mixed model. The significance tests of covariates are similar to the untransformed model (type I results shown in Table 5). Again, we see a significant time effect but no group effects. The change from baseline to 6 months is significant, $t(853) = -6.71, p < 0.0001$. The change from 6 to 18 months after baseline is not statistically significant, $t(842) = -0.32, p = 0.75$. As discussed above, the beta distribution models a variable that takes values on the interval $[0, 1]$, which is conceptually more appropriate for proportion data. The generalized linear mixed models can be fitted by *proc glimmix* in SAS, with appropriately specified distributions and link functions. The mixed-effects beta regression currently is not available in R, and other generalized linear mixed models such as mixed-effects logistic regression can be fitted by function *glmer* (“lme4”).

Summary

Proportion data are not necessarily non-normal and do not always require transformations or special analytic techniques. However, it is important to test the underlying statistical assumptions underlying planned analyses. For instances in which statistical assumptions are violated, it may be important to apply transformations or use generalized linear models in order to draw valid inferences about the data from the results. Beta regression is particularly well suited to non-binomial proportion scores, as when proportions of subscale scores are computed relative to the total score. These types of proportion scores are relatively common in prevention trials and warrant special attention.

This research was supported by a T32 Fellowship from the National Institute of Mental Health (NIMH) to the third author (T32MH018951) and by a grant from the NIMH to the fourth author (MH093508). The authors would also like to thank Dr. David Kolko for access to the SKIP2 (MH063272) dataset and Charles Bennett for the assistance with the manuscript preparation.

Compliance with Ethical Standards

Funding This research was supported by grants from the National Institute of Mental Health (MH093508; MH063272; MH018951).

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All data collection procedures were carried out with approval from, and in compliance with, the IRB at the University of Pittsburgh.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

- Alnosaier, W. S. (2007). *Kenward-Roger approximate F test for fixed effects in mixed linear models*. (Doctoral dissertation). Retrieved from <https://ir.library.oregonstate.edu/xmlui/handle/1957/5262>.
- Chamberlain, P., Price, J., Leve, L. D., Laurent, H., Landsverk, J. A., & Reid, J. B. (2008). Prevention of behavior problems for children in foster care: Outcomes and mediation effects. *Prevention Science, 9*, 17–27.
- Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software, 34*, 1–23. doi:10.18637/jss.v034.i02.
- Dobson, A. J. (2002). *An introduction to generalized linear models* (2nd ed.). London: Chapman and Hall.
- Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics, 31*, 799–815. doi:10.1080/0266476042000214501.
- Gotelli, N. J., & Ellison, A. M. (2004). *A primer of ecological statistics*. Sunderland: Sinauer Associates.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576. doi:10.1146/annurev.psych.58.110405.085530.
- Gregory, K. B. (2011). *A comparison of denominator degrees of freedom approximation methods in the unbalanced two-way factorial mixed model*. (Master’s thesis). Retrieved from <http://www.learningace.com/doc/2798578/695791137a52ad51d325eaa95f220bd1>.
- Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. New York: Springer.
- Kolko, D. J., Campo, J., Kilbourne, A. M., Hart, J., Sakolsky, D., & Wisniewski, S. (2014). Collaborative care outcomes for pediatric behavioral health problems: A cluster randomized trial. *Pediatrics, 133*, e981–e992. doi:10.1542/peds.2013-2516.
- Lindhiem, O., Shaffer, A., & Kolko, D. J. (2014). Quantifying discipline practices using absolute vs. relative frequencies: Clinical and research implications for child welfare. *Journal of Interpersonal Violence, 29*, 66–81. doi:10.1177/0886260513504650.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for mixed models* (2nd ed.). Cary: SAS Publishing.
- Liu, S., Rovine, M. J., & Molenaar, P. C. M. (2012). Selecting a linear mixed model for longitudinal data: Repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychological Methods, 17*, 15–30. doi:10.1037/a0026971.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.
- Olsson, U. (2002). *Generalized linear models: An applied approach*. Lund: Studentlitteratur.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Raim, A. M., & Neerchal, N. K. (2013). *Modeling overdispersion in binomial data with regression linked to a finite mixture probability of success*. *Proceedings of the Joint Statistical Meetings (ASA Section on Survey Research Methods)* (pp. 2760–2774). Alexandria: American Statistical Association.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–77. doi:10.1037/1082-989X.7.2.147.
- Schuhmann, E. M., Foote, R. C., Eyberg, S. M., Boggs, S. R., & Algina, J. (1998). Efficacy of parent–child interaction therapy: Interim

- report of a randomized trial with short-term maintenance. *Journal of Clinical Child Psychology*, 27, 34–45.
- Shaffer, A., Lindhiem, O., & Kolko, D. J. (this issue). Treatment effects of a primary-care intervention on parenting behaviors: It's all relative. *Prevention Science*.
- Shelton, K. K., Frick, P. J., & Wootton, J. M. (1996). Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*, 25(3), 317–329.
- Sokal, R. R., & Rohlf, F. J. (1995). *Biometry: The principles and practice of statistics in biological research* (3rd ed.). New York: W. H. Freeman.
- Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11, 54–71. doi:10.1037/1082-989X.11.1.54.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA*. Belmont: Duxbury.
- Warton, D. I., & Hui, F. K. C. (2011). The arcsine is asinine: The analysis of proportions in ecology. *Ecology*, 92, 3–10. doi:10.1890/10-0340.1.
- Wilson, K., & Hardy, I. C. W. (2002). Statistical analysis of sex ratios: An introduction. In I. C. W. Hardy (Ed.), *Sex ratios: Concepts and research methods* (pp. 48–92). New York: Cambridge University Press.
- Zar, J. H. (1998). *Biostatistical analysis* (4th ed.). Englewood Cliffs: Prentice Hall.