Multilevel random effect and marginal models for longitudinal data

Bedilu Alamirie Ejigu
Department of Statistics, Faculty of Natural and Computational Science,
Addis Ababa University, Ethiopia
Email: bedilue@gmail.com

ABSTRACT

In many clinical trials, in order to characterize the safety profile of a subject with a given treatment, multiple measurements are taken over time. Mostly, measurements taken from the same subject are not independent. Thus, in cases where the dependent variable is categorical, the use of logistic regression models assuming independence between observations taken from the same subject is not appropriate. In this paper, marginal and random effect models that take the correlation among measurements of the same subject into account were fitted and extensions on the existing models also proposed. The models were applied to data obtained from a phase-III clinical trial on a new meningococcal vaccine. The goal is to investigate whether children injected by the candidate vaccine have a lower or higher risk for the occurrence of specific adverse events than children injected with licensed vaccine, and if so, to quantify the difference. Moreover, in the paper, extensions for the random intercept partial proportional odds model and generalized ordered logit model which assumes identical variability for different category levels were extended by introducing category specific random terms. This is very appealing to study the association between different category levels. Instead of using the classical logistic regression, Generalized Estimating Equations (GEEs) and random effect models are appropriate when measurements taken from the same subject are not independent. The result reveals that, in both marginal and random effects model, significant difference between the two vaccines were found for pain and redness adverse event.

INTRODUCTION

Pharmaceutical companies develop vaccines which contains an agent that resembles a disease-causing microorganism in order to improve immunity to a particular disease. When a company aims to bring a new vaccine product to market, the safety profile of the vaccine is assessed in different ways to ensure that it is safe. In most cases, the clinical safety evaluation of the vaccine is performed regarding two specific aspects (Bergsma *et al.*, 2013). First, the occurrences of a certain number of local or general symptoms are checked proactively via diary cards recording the occurrence or absence of the symptom during a certain number of days after the injection. To properly assess the safety profile of the new vaccine, subjects injected with the vaccine are evaluated for different adverse event outcomes such as pain, redness and irritability over time. In most cases, for ease of recording a standard intensity scale that expresses the level of adverse event is often used and contains a certain number of possible intensity of the symptom. Subjects are then asked to fill in their maximum daily intensity of each reported solicited symptom during the entire solicited symptom follow-up period in the diary card. Based on such scales, one can then establish the vaccine and outcome relationship and test whether subjects injected by the candidate vaccine have a lower or higher risk for the adverse event than subjects injected by the licensed vaccine.

This paper will focus on the analysis of repeated categorical measurements concerning solicited symptoms coming from vaccine clinical trials. Specifically, the safety profile of a candidate vaccine for meningococ-

cal disease which is a life-threatening illness caused by strain of bacteria called Neisseria meningitides will be assessed.

Currently different vaccines such as Menactra, Menveo and Mencevax are available against Meningococcal infection (Food and Drug Administration (2014). The safety of the candidate vaccine for meningococcal disease is evaluated by comparing the level of redness, pain and irritability adverse events measured by ordinal scale at each follow up day to the one of a licensed vaccine. We will consider here a four-day follow-up period, the day of vaccination being denoted as day one and taken as a reference day in further analysis. Analysis methods presented hereafter will take the ordinal and correlated nature of the data due to repeated measures from the same subject using two model families: the marginal model family which is characterized by the specification of the mean function and the random effects family that focuses on the expectation conditional upon the random effects studied. The analysis will be done for primarily measured ordinal outcome as well as for the dichotomized outcome. Generalized estimating equations (Liang and Zeger, 1986) from marginal models are usually preferred to evaluate the overall adverse events as a function of treatment group and visiting day. While, in a random effects approach (Berslow and Clayton, 1993), the response rates are modeled as a function of covariates and parameters specific to a subject.

The two model families do not only differ in the questions they address, but also in the way they deal with the dependencies between the observations. This difference way of handling the within child association leads the two models families for different purpose as mentioned by different authors (Laird and Ware, 1982; Agresti, 2002; Fitzmaurice et al., 2004). As a result, interpretations of the regression model

parameters are different. For the partial proportional odds random intercepts model which assume identical baseline variability within the subject being in different categories of the outcome, extensions that allow to have different random effect variability at each category proposed.

MATERIALS AND METHODS
Case study: Phase-III clinical trial

The data used in this paper come from a phase III clinical trial evaluating the safety profile of a new vaccine for meningococcal infection. In the study, children with ages from 12 to 15 months are randomly assigned to the candidate and licensed vaccine with 3:1 ratio, respectively. Children after recruitment in the study were injected with the vaccine at the upper left thigh at day one, and the parents of the children were asked to fill in diary cards indicating whether or not their children experienced either pain, redness, or irritability during the follow-up period of four days (Bergsma et al., 2013). The levels of solicited symptom (Pain, Redness, or Irritability) were measured using ordinal scale. Pain and redness were measured only at injection site. Table 1 summarizes definition of solicited adverse event intensities. The data used in this study was collected from 1,880 children, of which 1,381 (73.5%) were assigned to the active treatment group (candidate vaccine for meningococcal) and the remaining 499 (26.5%) were assigned to the control group (licensed vaccine). Data from different children were assumed to be independent, but due to repeated measurements (of same child) over time, correlation is expected to exist. The primarily collected data consist of ordinal outcomes Y_{ij} for the observed solicited symptoms (Table 1) where Y_{ij} is the outcome for the ith child (i=1, 2, ... 1,880) at measurement day j (j =1, 2, .., 4).

Table 1. Variable Description

Adverse Event	Intensity	Description
Pain	0	Absent
	1	Minor reaction to touch
	2	Cries/protests on touch
	3	Cries when limb is moved/spontaneously painful
Redness	0	Absent
	1	Diameter of redness 10 mm
	2	Diameter of redness 10 - 30 mm
	3	Diameter of redness 30 mm
Irritability	0	Behavior as usual
	1	Crying more than usual/no effect on normal activities
	2	Crying more than usual/interferes with normal activities
	3	Crying that cannot be comforted/interferes with normal activities

In order to compare the effect of the treatment group at a certain level of intensity with the other level of intensity dichotomization will be done to the outcome variable Y_{ij} as follows.

- To model all observed symptom versus no symptom:

$$W_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is} \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

- Observing at least moderate intensity levels of the symptom versus less than moderate levels:

$$X_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is} \geq 2 \\ 0, & \text{otherwise} \end{cases}$$

- Severe intensity level of the adverse event versus lower than severe adverse event:

$$Z_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is} \geq 3 \\ 0, & \text{otherwise} \end{cases}$$

Statistical methodology

The types of model for data analysis highly depend on the nature, and measurement scale of the outcome variable. In this study, the level of measurement for the variable of interest is ordinal. Ordinal data are specific forms of categorical data, where the order of the categories is of importance. Models for binary data have been extended to ordinal categorical outcomes (Aitchison and Silvey, 1957; Genter and

Farewell, 1985; Agresti, 2002). Due to repeated measurements taken from the same child over time, observations cannot be considered as independent. Thus, in such cases, the use of classical logistic regression model assuming independence of observations taken from the same child may not be appropriate. In the subsequent Sections, appropriate marginal and random effect models that account for the correlated nature of the data will be presented

Marginal models for ordered categorical data

In the marginal models settings, the responses are modelled marginalized over all other responses (Molenberghs and Verbeke, 2005). Generalized estimating equations (GEE) introduced by Liang and Zeger (1986) is an intuitively appealing way to model longitudinal data in marginal models framework. The interest in standard GEE focuses on the relationship between the covariates and the probability of response while response correlation is treated as a nuisance parameter.

When the response categories are ordered, the use of this ordering yield more parsimoniously parameterized models. Further, the resulting odds ratios based on the dichotomized outcome may depend on the cut point chosen to dichotomize the outcome (McCullagh, 1980; Hosmer and Lemeshow, 2000). Models

that use cumulative probabilities like proportional odds models, adjacent categories logits and Continuation ratio logits (McCullagh, 1980; Ananth and Kleinbaum, 1997; Agresti, 2002) are possible choices for modeling ordinal data. Continuation-ratio model is suited when the underlying outcome is irreversible and adjacent-category model designed for situations in which the subject must 'pass through' one category to reach the next category (Liu and Agresti, 2005) are not used in this analysis. As a result, this paper will focus on ordinal logistic regression models under the GEE modeling framework.

Proportional odds model (POM)

The unique feature of proportional odds model (POM) is that the odds ratio for each predictor is taken to be constant across all possible collapsing of the response variable. When the assumption is met, odds ratios in a POM are interpreted as the odds of being lower or higher on the outcome variable across the entire range of the outcome (Scott *et al.*, 1997). In POM, reversing the direction of the response levels will change the direction of the effects but not their magnitude or significance (McCullagh, 1980; Hosmer and Lemeshow, 2000).

Let μ_{ijk} be the probability of the ith subject at the jth visiting day being in the response category k, $\mu_{ijk} = P(y_{ij} = k)$. Further, let the cumulative probability of the response in category k or above be represented by $\Pi_{ijk} = P(Y_{ij} \geq k)$. The lowest outcome which corresponds to a baseline level,

$$\Pi_{ij0} = \mu_{ij0} + \mu_{ij1} + ... + \mu_{ijk} = 1; \Pi_{ij1} = \mu_{ij1} + ... + \mu_{ijk}; \Pi_{ij2} = \mu_{ij2} + .. + \mu_{ijk}; \Pi_{ijk} = \mu_{ijk}.$$

The POM is represented as follows;

$$logit(\Pi_{ijk}) = \beta_{0k} + \beta_{1j}Day_{ij} + \beta_2 Trt_i \qquad (1)$$

where, k is the level of the ordered category. The parameter β_{0k} is the intercept for category k, usually considered as nuisance parameters of little interest (Agresti, 2002). Trt_i takes the value 1 when the

subject is assigned to the candidate vaccine and zero otherwise. Day_{ij} has four levels and equal to 1 when *response* observed at day j for subject i, zero otherwise (day one) taken as a reference day). The parameter **β**'s here have population averaged interpretations. Model (1) assumes an identical effect of the predictors for each cumulative probability. Specifically, the model implies that odds ratios for describing effects of explanatory variables on the response variable are the same for each of the possible ways of collapsing the response to a binary variable. Violation of this assumption leads to an incorrect model.

In GEE, estimates of the parameter **β** are obtained by solving the generalized estimating equations

$$S(\beta) = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \beta'} \left(A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}} \right)^{-1} (W_i - \mu_i) = 0 \qquad (2)$$

where A_i is a diagonal matrix with the marginal variance $(\mu_i) = Var(w_i)$ on the main diagonal and $R_i(\alpha)$ is the working a correlation matrix that depend on the unknown parameter vector α.

Liang and Zeger (1986) showed that using the method of moments concept, when the marginal mean has been correctly specified and when the mild regularity condition hold, the estimator obtained by solving the score equation (2) is consistent and asymptotically normally distributed with mean β and asymptotic variance covariance matrix var (β).

In summary, marginal models for longitudinal data separately model the mean response, and within child association among the repeated responses. The aim is to make inference about the mean response, whereas the association is regarded as nuisance characteristics of the data that must be accounted for to make valid inferences about changes in the population mean response. This separate specification of the mean and within child association has an important implication on parameter interpretation. Since the GEE approach does not specify completely the joint distribution,

likelihood-based methods to compare models and to conduct inferences about the parameter are not available. To draw inference in a quasi-likelihood approach, Boos (1992), Rotnitzky and Jewel (1990) illustrate a generalization of score tests for different models including models based on GEE.

Generalized linear mixed models

In many clinical/biomedical researches the longitudinal responses are not necessarily continuous. As a result, the general linear models and general linear mixed models might not apply. Thus, when the longitudinal responses are discrete, Generalized Linear Models (McCullagh and Nelder, 1989) are required to relate changes in the mean responses to covariates. Generalized linear Mixed Models (Berslow and Clayton, 1993) are obtained from GLMs by incorporating random effects into the linear predictors. Such random effect models can account for a variety of situations, including child heterogeneity, unobserved covariates and have conditional interpretation with child-specific effects (Liu and Agresti, 2005). The assumptions made in GLMMs are (i) conditional on the child-specific random effect (b_i) and covariates (Day_{ij}, Trt_i) the distribution of Y_{ij} belongs to exponential family (ii) the random effect b_i follows a normal distribution with a mean 0 and variance σ_{bi}^2 and (iii) conditional on b_i the repeated measures Y_{ij} are independent. In the context of this case-study, the generalized linear mixed model formulations for ordinal outcomes are presented in the following Subsections.

Random effect models for ordinal outcomes

The unique feature of proportional odds model (POM) is that the odds ratio for each predictor is taken to be constant across all possible collapsing of the response variable (Scott *et al.*, 1997). When proportional odds assumption is met and child

specific parameter estimates are of interest, partial ordinal model (POM) can be easily fitted in random effects modeling framework by introducing random effect terms (b_i) specific to child *i* in model (1). In this model, the ordinal nature of the response is taken into account by considering the cumulative probabilities, $1= \Pi_{ij0}=P(y_{ij}) \geq 0 \geq \Pi_{ij1}=P(y_{ij}) \geq 1 ... \geq \Pi_{ijk}=P(y_{ij}) \geq k$. The random effect POM is written as follows;

$$logit(\Pi_{ijk}) = \beta_{0k} + \beta_{1j} Day_{ij} + \beta_2 Trt_i + b_i \qquad (3)$$

where Π_{ijk} is the cumulative probability of the outcome ($Y_{ij} \geq k$) conditional upon other covariates, k is the level of the ordered category, β_{0k} is the intercept for category *k*, the parameters β_{1j} and β_2 represents conditional log-odds ratios of the grouped categories superior to the cutoff (*k*) compared to the categories inferior to *k*, and b_i is child specific parameter to the ith child. To relax the strong assumption of identical log-odds ratio for the outcome by the covariate association in POM, partial proportional odds model (PPOM) and generalized ordered logit model (GOLM) have been considered, and can be easily fitted using NLMIXED procedure in SAS.

Partial proportional odds model (PPOM)

When the proportional odds assumption applies to some but not all of the covariates, the partial proportional odds model (4) can be used. In this model only the effect of treatment allowed to vary across the category levels, while the effects of day fixed at each category as done in POM (Model 3).

$$logit(\Pi_{ijk}) = \beta_{0k} + \beta_{1j} Day_{ij} + \beta_{2k} Trt_i + b_i \qquad (4)$$

where, Π_{ijk} is the cumulative probability of the outcome ($Y_{ij} \geq k$) conditional upon other covariates, k is the level of the ordered category, β_{1j} is the

effect of the Day_{ij} at the kth category level of the outcome and β_{2k} measures the log odds effect of the treatment at k^{th} category level of the outcome given children in both treatment groups having identical covariate and random-intercept, b_i is child specific parameter to the ith child and β_{0k} is the intercept for category k. The covariates and random effects determine conditional mean Π_{ijk} and the regression coefficients β can be therefore interpreted as conditional effects of covariates (child-specific), given the random effects b_i. For instance, the parameter β_{2k} interpreted as the log odds ratio comparing a child injected with the candidate vaccine with another child injected by the licensed vaccine, both having identical covariate and random-intercept values on the kth category of the outcome.

Generalized ordered logit model (GOLM)

This general model permits to each covariate to have different effect at each category of the outcome. For the ordinal outcome variable Y_{ij} with predictors treatment group (Trt_i) and the measurement day (Day_{ij}), the cumulative log odds are modeled as follows:

$$logit(\Pi_{ijk}) = \beta_{0k} + \beta_{1jk}\,Day_{ij} + \beta_{2k}Trt_i + b_i \qquad (5)$$

where; Π_{ijk}, β_{2k}, β_{0k} and b_i have the same meaning as mentioned under model (4), now β_{1jk} has also category specific estimate like β_{2k}.

Model (4) is special cases of model (5) when the effect of day is similar at each category. The disadvantage of model (5) is the larger number of parameters to be estimated as compared with previous one. But, this higher in number of parameter cannot be considered as a disadvantage in a situation when model (5) better fits the data.

Relationship between marginal and random effect model parameters

Zeger *et al.* (1988) derived an approximate relationship for the population averaged parameters (from GEE) and subject specific parameters with random effect in the linear predictor given by:

$$\frac{\hat{\beta}^{RE}}{\hat{\beta}^{M}} = \sqrt{c^2\sigma_{bi}^2 + 1} \qquad (6)$$

where $\hat{\beta}^{RE}$ and $\hat{\beta}^{M}$ are parameter estimates based on random effect and marginal models, respectively. σ_{bi}^2 is the variance of the random intercepts and, $c^2 = 16\sqrt{3}/15\pi$. Hence from this relationship it is clear that conditional effects are usually larger than marginal effects, and increase as the variance σ_{bi}^2 increase.

The estimated standard deviation ($\widehat{\sigma_{bi}}$) for the random intercept model is used as a summary of heterogeneity for the study population. The estimated standard deviation ($\widehat{\sigma_{bi}}$) equal to zero implies that the random intercept proportional logistic model (3) simplifies to a simple proportional logistic regression model treating all observations as independent. The size of estimated variance σ_{bi}^2 used to determine the scale on which the fixed effects should be judged. Moreover, the random part can be interpreted using measures of dependence. This is due to the fact that, unobserved heterogeneity between subject induces within subject dependence. Thus, in logit random intercept model, correlation (ρ) of the latent responses at any two occasion i and j given by $\rho = \sigma_{bi}^2/(\sigma_{bi}^2 + \pi^2/3)$ (Fitzmaurice *et al.*, 2009).

Proposed Extensions

Random effect models for ordinal outcome proposed by Harville and Mee (1984), Jansen (1990), Ezzet and Whitehead (1991) are restricted only to one random effect for all category. The

question is why we assume only one and similar random effect for different categories? To overcome such problems, we extend models (3) and (4) which assume identical baseline variability within the child being in either of the categories, by allowing to have different random effects b_{ik} at each category.

$$logit(\Pi_{ijk}) = \beta_{0k} + \beta_{1jk} Day_{ij} + \beta_{2k}Trt_i + b_{ik} \qquad (7)$$

where b_{ik} is the random intercepts for each category of the model and the vector of these random effects assumed to follow a multivariate normal distribution with mean vector zero and (co)variance matrix **D** (i.e. $b_{ik} \sim N(\mathbf{0}, \mathbf{D})$).

$$\mathbf{D} = \begin{pmatrix} d_{11} & \cdots & d_{1k} \\ \vdots & \ddots & \vdots \\ d_{k1} & \cdots & d_{kk} \end{pmatrix} \qquad (8)$$

where **D** is k x k general covariance matrix with elements d_{rs}. The elements of the matrix, d_{rs} represents the (co)variance between b_{ir} and b_{is} (r=1, 2, 3; s=1, 2, 3).

The advantage of the extended model over model (3) and (4) is that, it enables us to study the association between different category levels using b_{ik} covariance matrix. All the considered random-effects models are fitted by maximization of the marginal likelihood, obtained by integrating out the random effects. Since the likelihood function does not have a closed form in this case, model fitting is not an easy task. Numerical approximations will be used (Molenberghs and Verbeke, 2005) to maximize the marginal likelihood. In GLMMs, although in practice one is usually primarily interested in estimating the parameters in the marginal distribution for Y_{ij}, it is often useful to calculate estimates for the random effects b_i as well. They reflect between-child variability, which makes them helpful for detecting special profiles (i.e. outlying individuals). Moreover, estimates for the random effects are needed whenever interest is in prediction of child-specific evolutions.

RESULTS OF THE CASE STUDY

Recall that the aim of the study is to test whether there is a difference between the candidate and licensed vaccine in terms of percentage of children's who showed any level of solicited symptom (Pain, Redness, Irritability) taking into account the correlated nature of the data, and if there is a difference, to describe how does this difference between treatment groups develops over time.

Marginal models

When the responses are ordinal the usual test of independence ignores the ordering information to test whether there is association between the response and treatment. To test the general association between the two treatment groups and their response at the end of the study, a general Cochran-Mantel-Haenszel statistics had been performed and significant (p-value=0.001) result was obtained. The assumption of proportional odds (identical log odds ratios across different categories) was tested using score test and significant result was observed ($x^2_{(8)}$= 34.07, p-value=0.0001). Furthermore, from a separate analysis, it was found that the effect of the candidate vaccine on the outcome vary across different categories (Table 2) which is an indication for the violation of proportional odds assumption.

Therefore, drawing valid inference based on parameter estimates obtained by fitting model (1) may not be valid. For instance, based on POM, subjects injected with the licensed vaccine are 1.30 (1/0.77) times more likely to show pain adverse event than subjects injected with the candidate vaccine controlling other factors. While, based on a separate analysis, this odds ratio increases to 1.56 (1/0.64) and 3.03 (1/0.33) to observe at least moderate and severe intensity levels of pain, respectively (Table 2). Since the procedure PROC GENMOD in the current

Table 2. Estimated odds ratio for the candidate vaccine at different level of intensity.

	Modeled Level of Intensity							
	All		Moderate		Severe		POM	
Adverse Event	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI
Pain	0.86	(0.72,1.02)	0.64	(0.52,0.80)	0.33	(0.21,0.54)	0.77	(0.65,0.91)
Redness	0.77	(0.66,0.91)	0.70	(0.54,0.91)	0.96	(0.58,1.59)	0.75	(0.64,0.88)
Irritability	0.86	(0.72,1.02)	0.78	(0.58,1.03)	0.79	(0.46,1.36)	0.85	(0.72,1.02)

version of SAS 9.2.1 does not allow to deal with partial POM and GOLM in the marginal modelling framework, model (4) and model (5) will be fitted using NLMIXED procedure in SAS. Hence, analysis for a more general ordinal model that allow the effect of the vaccine to vary across different category levels as needed will be done under Section 3.2 in the random effects model framework.
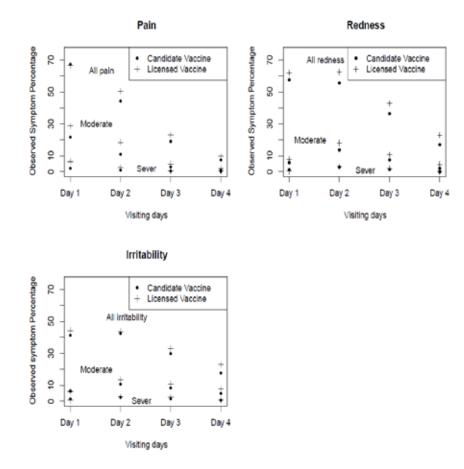

Figure 1. Percentage of solicited symptoms by treatment group at each visiting day

Generalized random effect models

In this section results based on an alternative approach using child (cluster) level terms in the model will be discussed. Given the discrete nature of time varying covariate (day) a random intercept model which adjusts only the intercept but does not modify the fixed effects was considered.

Random effect models for ordinal outcomes

The assumption of proportional odds across different categories was tested using the likelihood ratio test, by comparing model (3) with model (4) and model (5). In line with the marginal model results, the test suggests that, proportional odds assumption was not satisfied (p-value=0.0015). Therefore, partial proportional odds model (4) and generalized ordered logit model (5) were fitted for the ordered outcomes. The estimated OR's with their respective 95% CI obtained by fitting model (3), model (4) and model (5) are summarized in Table 3. The results showed that,

when we fitted the extended model (7) in the case of PPOM using category specific random effect terms, over model (4) which assumes only the same baseline variability at each category, the difference between the two treatment groups increase for high cutoff points (Table 3).

For instance, based on model (4), the odds ratio comparing a child injected with the candidate vaccine with another child injected by the licensed vaccine, both having identical covariate and random-intercept values at 3rd category of the outcome is 0.329, while based on model (7) it is 0.065. The results for the fixed effect parameters from the three models generally agree in terms of indicating children injected with candidate vaccine are less likely to show at least moderate and severe intensity levels of pain as indicated by less than one odds ratio (Table 3). Since, the 95% CI does not include one for the effect of treatment at category one and three, this differences between

Table 3. Maximum Likelihood Estimates and Approximate 95% Confidence Intervals for PPOM based on model (4), and model (5) and GOLM with random intercept for pain adverse event

Parameters	Model (4)		Model (5)		Model (7) for PPOM	
	OR	(95%CI)	OR	(95%CI)	OR	(95%CI)
Fixed Effect: Odds Ratios						
β_{21} Treatment at category 1	0.750	(0.561, 1.003)	0.751	(0.562, 1.001)	0.765	(0.545,1.073)
β_{22} Treatment at category 2	0.513	(0.367, 0.719)	0.523	(0.371, 0.732)	0.401	(0.239,0.672)
β_{23} Treatment at category 3	0.329	(0.188, 0.576)	0.344	(0.198, 0.593)	0.065	(0.015,0.292)
Random Effects: Variances						
$Var(b_i)$		5.46		5.55		
$Var(b_{i1})$						7.618
$Var(b_{i2})$						12.710
$Var(b_{i3})$						40.468

Treatment at category k (k =1,2,3) refers OR of the candidate vaccine on outcome category k

the two treatment groups are statistically significant at 5% level of significance. Approximately, similar residual intra-class correlation was obtained for the latent responses and was estimated as 0.63 ($\sigma_{bi}^2 / (\pi^2/3 + \sigma_{bi}^2)$) for both model (4) and model (5).

Extensions for random effect models with ordinal outcomes

Even though, it is very computationally intensive due to the increased number of random effects, to allow different category specific random terms and to study the association between category levels, the extended model (7) is fitted for pain adverse event. A significant (p-value=0.001) positive association (ρ_{12}= 0.33) between the first and the second category level random effects was observed from the estimated (co) variance matrix. This suggests that the correlation between the first category (all intensity levels of pain) and the second category (at least moderate intensity levels of pain) for child i at visiting day j less or equal to the correlation between the two random effects ($\rho_{12} = 0.33$). While, since the third random effect term is independent (with p-value ≥ 0.39) with the first and the second category level random effects term, significant association between severe intensity level with the other category levels were not observed. The estimated random effects (co) variance matrix which reflects the baseline heterogeneity between children at each category of the outcome is given by the following matrix (* refers significantly different from zero at 5% level of significance).

$$D = \begin{pmatrix} 7.618 * & 5.69 * & -1.495 \\ & 12.71 * & -0.33 \\ & & 40.46 * \end{pmatrix}$$

The variability between children at severe intensity level of pain (δ_{bi3}^2=40.468) is higher as compared with the variability observed in at least moderate (δ_{bi2}^2=12.71) and all intensity levels of pain (δ_{bi1}^2 =7.618).

This relatively large baseline heterogeneity between children implies that, there is substantial category specific variability in the propensity to experience a certain levels of adverse event. For instance, approximately 95% of the children have a baseline risk of showing at least moderate levels and severe level of pain symptoms that vary from (0.007% to 98.83%) and (0% to 99.86%), respectively. Based on the estimated category specific variance of the random intercepts, the estimated intraclass correlation for the latent responses becomes 0.70, 0.79 and 0.93 at category one (all), category two (at least moderate) and category 3 (severe), respectively.

DISCUSSION

In this paper, marginal and random effect models were used to analyze repeated categorical measurements concerning solicited symptoms coming from vaccine clinical trials. In case of marginal models the correlated nature of the data is acknowledged inside the estimating equation, while for random effects model it is done through the random effect part. Even though, both the considered marginal and random effect models are extensions of generalized linear model (McCullagh and Nelder, 1989), the different way of accounting within child association has a consequence for the interpretation of the regression model parameters. In the random effect approach the goal is to determine child specific changes in the risk of observing solicited symptoms over the courses of the study, while in the marginal model the emphasis is to determine the overall change.

To fix ideas, let us reconsider the estimated effect of the candidate vaccine under different random effect model formulation (Table 4).

The estimated treatment effect 0.449 and 0.162 from the random effect model describes how the odds of observing at least moderate and severe levels of pain

Table 4. Summary of estimated odds ratios (95%CI) of the candidate vaccine based on random effects model
for the three adverse events

Fitted Model	Adverse Event	Modeled intensity level of adverse event			AIC
		All	Moderate	Severe	
Model (4)	Pain	0.75(0.56,1.00)	0.51(0.37,0.72)	0.33(0.19,0.58)	10266
	Redness	0.61(0.46,0.79)	0.61(0.43,0.85)	1.22(0.68,2.18)	11472
	Irritability	0.76(0.53,1.09)	0.78(0.51,1.22)	0.80(0.41,1.53)	10355
Model (5)	Pain	0.75(0.56,1.00)	0.52(0.37,0.73)	0.34(0.20,0.59)	10261
	Redness	0.60 (0.45,0.79)	0.61(0.43,0.86)	1.21(0.67,2.17)	11334
	Irritability	0.76(0.52,1.09)	0.79(0.51,1.23)	0.80(0.42,1.53)	10270
Model^a(7)	Pain	0.77(0.55,1.07)	0.40(0.24,0.67)	0.07(0.02,0.29)	10087
Model^b (7)	pain	0.77(0.56,1.07)	0.38(0.22,0.68)	<0.01(0.00,0.003)	10029

AIC=Aikakes Information Criterion, *Model^a* (7)=non-proportional odds only for treatment, *Model^b* (7)=non-proportional odds for both treatment and visiting day

increase for any child treated with the candidate vaccine. Note that, the same odds ratio is significant across both the random effects model and marginal models, but the magnitude of the effect can differ.

Therefore, the answer for the question "how the candidate vaccine is beneficial?" will depend on whether the interest is in its impact on the study population or on an individual drawn from that population. The estimated treatment effect difference from the marginal model somewhat smaller than the estimated treatment effect from random effect model (equation 6), and this discrepancy becomes high when the estimated variance for the random intercept based on the random effect models becomes large. These differences in the estimated coefficients and odds ratios are due to different interpretations of in the two model families, which are these two classes of models estimate parameters that address substantively different questions. More precisely, the estimates of fixed effects of the treatment in model (5) describe the effect of the treatment on a specific child to observe solicited symptom. While, the corresponding effects in the marginal model (1) describe the effects of treatment on the prevalence of observing solicited symptom in the population of children injected by the candidate vaccine. The approximate relationship between the two model parameters mentioned in Section 2.3 highlights how the parameters estimates for marginal model are attenuated relative to the corresponding fixed effects in model (5).

If the interest is to model the heterogeneity among children and to draw likelihood based inferences, we prefer to fit random effect models over GEE. In random effects model, each child is assumed to have its own level of adverse event. Thus, it is well known that fixed effects parameters do not maintain their interpretation when random effects are introduced in the model. Therefore the fixed effects odds ratio no longer is an odds ratio between any two children as mentioned by Zeger *et al.* (1988).

Among the considered models that account the ordinal nature of the data, the extended model (7) with category specific random terms better fits the data (Table 5) for pain adverse event.

Table 5. Summary of Goodness of fit measures for the fitted ordinal random effect models for pain adverse event

Measure	One Random effect			Based on extended models (Three random effect)		
	POM	PPOM	GOLM	POM	PPOM	GOLM
-2loglik	10250	10246	10229	10069	10057	10029
No. Parameters	8	10	16	13	15	21
AIC	10275	10266	10261	10095	10087	10071
R^2_{ms}	0.430	0.431	0.433	0.450	0.450	0.453

Further, since the considered random effects model accounts only baseline heterogeneity, it is also possible to extend those models by introducing visiting day specific random effects b_{ij}, where b_{ij} is the random effect for subject i at occasion day j (j =1 to 4) and vector of b_{ij} follows a multivariate normal distribution with mean vector 0 and (co) variance matrix D. But, in practice, when the number of visiting days increases, it is difficult to introduce more than few random effects due to computational intensive integration methods. Even the result presented in Table 4 for model[a](7) and model[b](7) with three random effect terms took approximately 470:18 hour and 774:37 hour on 2.5GHz PC, respectively. Moreover, NLMIXED procedure fails to integrate the likelihood for less than 20 quadrature points. Among the considered optimization techniques (Levenberg-Marquardt Method, Newton-Raphson, Trust-Region Method), all the presented parameter estimates were obtained using Newton-Raphson optimization techniques.

CONCLUSION

Both marginal and random effect modeling approaches provide similar conclusions about the significant difference between the two vaccines. A significance difference between the candidate and licensed vaccine in terms of percentage of children who showed at least moderate and severe intensity levels of pain, all and at least moderate intensity levels of redness were found in both modeling approaches. The difference between the two treatment groups become high for severe intensity level of pain as compared with the other intensity levels of solicited symptoms. In both marginal and random effect models, significant difference between the two treatment groups were not found for all intensity levels of pain symptom and severe intensity level of redness. Further, in all analyses, significant difference was not observed between the two treatment groups for all, at least moderate and severe intensity levels irritability. Moreover, non-significant interaction effect in all the considered adverse event models implies that, the difference between the two treatment groups can be considered constant over the follow-up period. In addition to this, the extended models better fits the data than the existing methods.

REFERENCES

Agresti, A and Liu, I. M. (1999). Modeling a categorical variable allowing arbitrarily many category choices. *Biometrics* **55**:936-943.

Agresti, A. (2002). Categorical Data Analysis, (2nd edition). New York: Wiley.

Aitchison, J and Silvey, S. D. (1957). The generalization of probit analysis to the case of multiple responses. *Biometrika* **44**:131-140.

Ananth, C.V and Kleinbaum, D. G. (1997). Regression models for ordinal responses: A review of methods and applications. *International journal of Epidemiology* **26**:1323-1333.

Bergsma, W.P., Aris, E. M. D and Tibaldi, F.S. (2013). Linear categorical marginal modeling of solicited symptoms in vaccine clinical trials. *Statistics in Biopharmaceutical Research* **5:** 27-37.

Boos, D. (1992). On Generalized score tests. *American Statistician* **46**: 327-333.

Diggle, P. J., Heagerty, P. J., Liang, K. Y and Zeger, S. L. (2002). *Analysis of Longitudinal Data* (2nd edition). Oxford: Oxford University Press.

Fitzmaurice, G. M. (1995). A caveat concerning independence estimating equations with multiple multivariate binary data. *Biometrics* **51**: 309-317.

Fitzmaurice, G. M., Laird, N. M and Ware, J.H. (2004). Applied Longitudinal Analysis. John Wileyand Sons, Inc., Hoboken, New Jersey.

Fitzmaurice, G., Davidian, M., Verbeke, V and Molenberghs, M. (2009)**.** Longitudinal Data Analysis**:** Handbooks of Modern Statistical Methods. Chapman and Hall/CRC

Food and Drug Administration (2014). www.**fda. gov/.../BiologicsBloodVaccines/Vaccines/.../ UCM201349.pdf**.

Frees, E.W. (2004). Longitudinal and Panel Data: Analysis and Application in the Social Sciences. Cambridge University Press.

Genter, F. C and Farewell, V. T. (1985). Goodness-of-link testing in ordinal regression models. *Canada Journal of Statistics* **13**: 37-44.

Klaus Larsen, K., Petersen, H.J., Jorgensen, B.E and Endahl, L. (2000). Interpreting parameters in the logistic regression model with random effects. *Biometrics* **56:** 909-914.

Laird, N. M and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**: 963-974.

Lin, K. C. (2010). Goodness of fit tests for modeling longitudinal ordinal data. *Computational Statistics and Data Analysis* **54:**1872-1880.

Liu, I and Agresti, A. (2005). The Analysis of Ordered Categorical Data: An overview and a survey of recent developments. *Sociedad de Estadistica e Investigacion Operativa Test.* **14:**1-73.

McCullagh, P and Nelder, J.A. (1989). Generalized Linear Models, 2nd Edition, and London: Chapman and Hall.

Molenberghs, G and Verbeke, G. (2005). Models for Discrete Longitudinal Data. New York: Springer-Verlag.

Nick, R.P., Matthew, L.C., Juul, A and Nigel, S. (2009). Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R. *Computational Statistics and Data Analysis* **53:**632-641.

Pollard, A. J and Maiden, M. C. J. (2001). *Meningococcal Vaccines: Methods and protocols*. Humana Press.

Rotnitzky, A and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semi-parametric generalized linear models for cluster correlated data. *Biometrika* **77:**485-497.

SAS Institute Inc. 2008. SAS/STAT 9.2 Users Guide. Cary, NC: SAS Institute Inc.

Scott, S.C., Goldberg, M.S and May, N. E. (1997). Statistical assessment of ordinal outcomes in comparative studies. *Journal of Clinical Epidemiology* **50:**45-55.

Verbeke, G. and Molenbergs, G. (2000). Linear Mixed Models for Longitudinal data. Springer-Verlag, Berlin.