Index of Lecture 10b: Logistic Mixed Models and Alternative Methods

Practical Information

Today's session:

- mixed models for discrete data[1], with focus on random effects logistic regression,
  - illustrated by two-level pig dataset (`pig_adg`) previously explored,
- "alternative methods" — to the linear and logistic mixed models:[2]
  - perhaps more relevant for discrete than continuous data[3],
  - brief review of several methods, contrasting their properties,
  - main focus on generalized estimating equations (GEE), but really just an introduction to a large topic,
  - illustrations and comparisons of methods by familiar 2-level datasets: simulated datasets (Chapter 20) and `pig_adg` data.

News/Schedule:

- home assignment 4 and project proposal due on Monday (March 17),
- two VER problems posted, to work with for Monday's lab session.

---

[1] Corresponding to Sections 22.1-4 of the VER2/MER textbook (with some skippings).
[2] VER2/MER textbook coverage: alternative methods: Section 20.5, GEE: Section 23.5.
[3] Linear mixed models remain the most popular choice when its assumptions can be met.

Mixed Models for Discrete Data

Synthesis:
discrete mixed models — usually called generalized linear mixed models (GLMMs) — extend and combine linear mixed models (LMMs) and generalized linear models (GLMs),

- intuitive and flexible modelling of clustering by means of random effects,
- but interpretation of models are less straightforward than in LMMs and GLMs,
  - fixed effects come in two versions: cluster-specific and population-averaged,
  - variance components more difficult to estimate meaningfully (and maybe variances are less meaningful for discrete data anyway),
- and statistical analysis of GLMMs is not as simple as we would want/are used to: many methods exist[4] and none has proven generally superior or always useful.

Contents:

- very brief introduction to random effects logistic regression model,
- overview of statistical inference (but limited details),
- brief Stata demonstration (note: models not available in Minitab).

[4] Referring to methods for estimation and inference; alternative modelling approaches exist as well (later in lecture).

Random Effects in Logistic Regression

Logistic regression: the linear modelling of x's is at logit scale.

Idea: put random effects also at logit scale.

Somatic cell count example (conceptual continuation):

- Let Y_{ij} = binary disease status wrt. clinical mastitis for cow i in herd j,
- let $p_{ij} = \Pr(Y_{ij} = 1)$ (disease probability),
- model with random herd effects (u_j):

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_{1ij} + \ldots + \beta_k x_{kij} + u_j, \qquad (1)$$

  where $u_j \sim N(0, \sigma_h^2)$, and σ_h is the dispersion (standard deviation) among herds in logit(disease probabilities):

  - $\sigma_h = 0$: no disease clustering in herds (after accounting for fixed effects),
  - σ_h large: strong disease clustering within herds (obs. in a herd are very alike), and large variation between herds (obs. are very different between herds).
  - value of σ_h has no simple interpretation for p's.

Note: Equation (1) is conditional on the u_j's and gives the probability of disease for fixed u_j's[5]; animals within a herd are also assumed conditionally independent.

[5] To compute p_{ij} when the herd random effect is unknown (a marginal probability) is difficult and can only be done numerically; Stata's `margins` command will do this after some model estimations, see 10bL−22.

Random Effects in Generalized Linear Models

We build a GLMM from a GLM in the same way as with the random effects logistic regression:

- put the random effects at the transformed scale (determined by the link function), for example,
  - binary regression with link function g (logit or alternatives):
$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + u_{\text{herd}(i)},$$
  - Poisson regression with log link:
$$\ln(\lambda_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + u_{\text{herd}(i)},$$
  - general GLMM formulation:
$$g(\mu) = \eta = X'\beta + Zu, \tag{2}$$
    where Z is the design matrix for the random effects,
    - more than one set of random effects is possible, e.g. for a 3-level hierarchy,
    - the notation Zu is just short for all random effects in the model,
- models with a cluster-specific (as opposed to population-averaged) interpretation of the fixed effects parameters, see 10bL−7,
- no simple variance decomposition (as we had for linear mixed models); see 10bL−8 for logistic model, and textbook Chapter 22 for Poisson model.

Pig Data Example

We revisit the pig_adg dataset (Model-building exercise), and occurrence of enzootic pneumonia and its dependence on atrophic rhinitis (dichotomized at ar > 1, ar_g1).

Data structure:

Level	Number	Mean	Range
Farm	15	–	–
Pig	341	22.7	14–28

2×2 table analysis:

simple OR $= 1.909$
ln(OR) $= 0.647$

	pn		
ar_g1	1	0	Total
1	109	66	175
0	77	89	166
Total	186	155	341

Simple logistic regression for pn:

$$p_{ij} = \Pr(\texttt{pn}_{ij} = 1), \quad \text{for pig } i \text{ in herd } j,$$
$$\text{logit}(p_{ij}) = \beta_0 + \beta_1\, \texttt{ar\_g1}_{ij},$$

gives $\hat{\beta}_0 = -0.145$ (SE$=$0.258) and $\hat{\beta}_1 = 0.647$ (SE$=$0.220).

Expand to a **random effects** logistic regression to account for herds:

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1\, \texttt{ar\_g1}_{ij} + u_j,$$

where u_j's are the random herd effects, assumed $\sim \text{N}(0, \sigma_h^2)$.

- $\hat{\beta}_1 = 0.437$ (SE$=$0.258) — much smaller than above, probably due to confounding by herds, and also much weaker ($P = 0.091$ compared to $P = 0.003$),
- $\hat{\sigma}_h^2 = 0.877$ (SE$=$0.433) — some herd clustering and clearly significant,
- $\hat{\beta}_0 = 0.020$ (SE$=$0.301) — larger SE because now a herd-level parameter.

Reunion Island Data Example

Binary outcome = conception at first service (yes/no):

- observed in each lactation of the participating cows,
- 3-level data structure (disregarding again the 5 regions at the top level):
  3027 lactations — 1575 cows — 50 herds,
- overall conception rate: $1302/3027 = 0.43$,
- fixed effects: `heifer`, `ai` and log calving to first service interval (`lncfs`).

Random effects logistic regression:

$$\text{logit}(p_{ijk}) = \beta_0 + \beta_1\,\texttt{heifer}_{ijk} + \beta_2\,\texttt{ai}_{ijk} + \beta_3\,\texttt{lncfs}_{ijk} + v_k + u_{jk}, \qquad \text{where}$$

- p_i = prob. of conception in lactation i of cow j in herd k,
- $v_k \sim N(0, \sigma_h^2)$ and $u_{jk} \sim N(0, \sigma_c^2)$.

Parameter estimates (from Stata's `melogit` command):

$$\text{intercept}: \ \hat{\beta}_0 = -1.515\,(\text{SE}=.446); \quad \texttt{heifer}\ \text{coef.}: \hat{\beta}_1 = -0.062\,(\text{SE}=.097),$$

$$\texttt{ai}\ \text{coef.}: \ \hat{\beta}_2 = -1.023\,(\text{SE}=.131); \quad \texttt{lncfs}\ \text{coef.}: \hat{\beta}_3 = 0.497\,(\text{SE}=.102),$$

$$\text{herd variance}: \ \hat{\sigma}_h^2 = 0.073\,(\text{SE}=.036); \quad \text{cow variance}: \hat{\sigma}_c^2 = 0.280\,(\text{SE}=.122).$$

— very small random effects...

Cluster-Specific versus Population-Averaged Interpretations

Odds-ratio (OR) computed as usual: OR $= \exp(\hat{\beta}_r)$: e.g., $\exp(0.44) = 1.55$ for `ar_g1`.

This OR has a cluster-specific interpretation:

- it refers to the difference in risk when comparing two pigs within the *same farm*, with/without the `ar_g1` condition,
- no assumptions are made about the actual farm (its random effect),
- "cluster-specific" ∼ for a specific farm; also (traditionally) denoted "subject-specific"[6], but here it does not refer to individuals!,
- the odds-ratio can be thought of as the value relevant to the farmer.

Alternative[7]: a population-averaged (or marginal) interpretation (directly from some statistical procedures, or by the approximation formula: $\beta_{\text{PA}} \approx \beta_{\text{CS}} / \sqrt{1 + 0.346\,\sigma_h^2}$):

- it refers to the difference in risk when comparing two pigs, with/without the `ar_g1` condition, from the *entire population* of pigs (not from the same farm),
- "population-averaged" ∼ across the population (of farms),
- the odds-ratio can be thought of as the value relevant to the association of farmers (or the slaughterhouse).

---

[6] The term "subject-specific" stems from repeated measures data, where individuals can be thought of as clusters; VHM 802 Session 12.

[7] Cluster-specific and population-averaged estimates agree generally in linear mixed models, and also in GLMMs when clustering is absent.

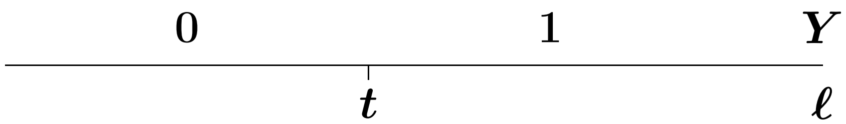
Variance Components in GLMMs

Interpretation of random effect variances:

- strictly speaking, as the variation in logit(p_i), or $g(\mu_i)$, between farms (clusters) after having accounted for fixed effects,
- no simple variance decomposition in GLMMs, because the model has
  - no error term (ε) and hence no σ^2 in the model equation,[8]
  - the total variance depending on the mean (and therefore not constant!).

Approximate method for random effects logistic regression[9]:

- put $\sigma^2 = \pi^2/3 = 3.29$, and proceed as for linear mixed models,
- gives approximate variances/correlations (albeit constant!) for the binary outcome Y,
- gives exact variances/correlations for a continuous latent (unobserved) variable ℓ associated with every cow, $\ell \sim$ "degree of sickness" (on some scale),
  - binary outcome Y corresponds to thresholding of underlying ℓ at cut-off value t:



formally, $Y = \begin{cases} 1, & \ell \geq t, \\ 0, & \ell < t, \end{cases}$

[8] The variation is on two different scales: observation scale and logit scale (σ_h^2).
[9] Reference for this/other methods: Vigre et al. (2004), *Preventive Veterinary Medicine* 63, 9−28.

Statistical Inference for GLMMs

Estimation is not straightforward[10]:

- different methods exist; our focus here is on maximum likelihood estimation in Stata,[11]
  - considered among best methods whenever feasible (Bayesian estimation also good, but beyond our current scope),
  - numerically difficult for large datasets/models,
  - cautious modelling recommended, for example by comparison with other approaches (e.g. GEE, later slides).

Tests and confidence intervals (approximate, even if model assumptions are met)

- fixed effects: Wald test and CI (the usual way), likelihood-ratio test preferable/necessary in the presence of estimation problems (e.g. extreme fixed effects estimates),
- random effects: Wald test and CI very crude, preferred test is likelihood-ratio test (if any testing is needed. . .),
- model checking: involves multiple levels as in linear mixed models,
  - lowest level suffers from same limitations as in ordinary logistic regression[12]
  - less statistics and procedures available, and options strongly software dependent.

[10] The main problem is that the likelihood function is difficult to compute, and can only be approximated.
[11] ML-estimation in Stata for GLMMs is superior to all other general statistics packages; in other software, one should check what approximation is used; some are of so poor quality that log L values are pretty useless for inference.
[12] With the additional random effects, covariate patterns with sensible replication are even harder to get by.

Overview: Dealing with Clustering

Detection of clustering:

- primarily through understanding of data structure,
- decisions about accounting for clustering should not be based on statistical testing![13]

Approaches to model clustering (in the course):

- mixed (random effects) models — covered already,

Approaches to account for clustering (in the course):

- fixed effects,
- stratification, for binary data (Mantel-Haenszel procedure),
- robust standard errors,
- generalized estimating equations (GEE).

Choice of approach should take into account:

- ease of use (but easy does not mean valid. . .),
- the assumptions required,
- (degree of) interest in clustering per se.

[13] The reason is that even small (possibly statistically non-significant) clustering can have substantial impact (as discussed in terms of variance inflation), so we prefer to account for clustering whenever its presence makes biological sense.

Fixed Effects Modelling

Fixed effects modelling: enter the herds (clusters) into the model as a categorical variable, to estimate a separate parameter for each herd (but one).

Advantages and disadvantages of fixed effects modelling:

- \+ generally very easy to carry out (e.g., a 2-level model will have no random effects),
- \+ avoids distributional assumption about herd effects,
- \+ avoids taking the herds as representative for a population (which might be inappropriate if the number of herds is small),
- +/− estimates are cluster-specific and specific to actual herds in the study (cannot be generalized to a population),
- − does not allow for herd-level predictors,
- − does not give an estimate of the variance between herds,
- − may lead to biased estimates for other fixed effects when the number of herds is large, in particular for non-normal models.

Results for simulated datasets (cow-level predictor only):[14]

- linear model: $\hat{\beta} = 4.968\,(.149)$ – close to mixed model,
- logistic model: $\hat{\beta} = 0.704\,(.046)$ – close to mixed model.

[14] Estimates for the intercept ∼ reference herd ⇒ not comparable to mixed models or models without herd effects.

STRATIFICATION (BINARY DATA)

Stratification = Mantel-Haenszel procedures using herds (clusters) as strata:

– combined odds-ratio (OR) across herds (binary within-herd predictor; binary outcome),

– test for association in two-way table (categorical within-herd predictor; categorical outcome), adjusted for herds.

Advantages and disadvantages of stratification:

+ easy to carry out,

+ avoids any assumptions about the herds,

+/− cluster-specific estimates (OR),

− restricted scope and limited analysis,

− no insight into the type or magnitude of clustering.

Results for simulated dataset (cow-level predictor only):

- Mantel-Haenszel OR $= 2.009 \Rightarrow \hat{\beta} = \ln(2.009) = 0.698$ − close to mixed model,
- estimated SE = .046, also close to mixed model (computed by backtransforming OR and its confidence interval bounds to logit scale[15]).

[15] The CI for the M-H OR was constructed on logit scale, where estimates follow a normal distribution to a good approximation, and transformed to odds-ratio scale.

Robust Standard Errors

Background:
usual (model-based) statistical methods:

$$\text{data} \mapsto \text{model} \mapsto \begin{cases} \text{estimates, test statistics} \\ \text{standard errors, } \boldsymbol{P}\text{-values} \end{cases}$$

- statistical models are not always (never) true!
- robust methods are designed to be less sensitive to model deviations.

Robust variance estimation (Huber-White, "sandwich"):

- base SEs on properties of the estimation method valid for a wider class of models than assumed,
- method can be targeted to deal with clustering, where the assumption of within-group independence is critical (only this version is of interest here).

Advantages and disadvantages of robust SE method:

+ simple to use and general method (available in Stata for wide range of models),

+/− robust SEs have a different interpretation than usual SEs,

+/− does not affect the estimate, only its standard error,

− gives no insight into the type or magnitude of clustering.

Results: fairly close to mixed model SEs (except: linear model, herd-level $\boldsymbol{X}$).

Generalized Estimating Equations (GEE)

Initial remarks about GEE for GLMs[16]:

- an estimation procedure (set of equations from which estimates are constructed iteratively) rather than a model,
- partially specified model involving only assumptions about the marginal[17] means and variances,
- gives population-averaged (or marginal) estimates,
- no herd effects, hence no assumed distribution for them,
- no likelihood function or likelihood-based inference.

Original form of GEE for GLMs:

- framework of longitudinal data (repeated measures),
- algorithm estimates a working correlation matrix for the correlation of observations within clusters (herds, subjects):
  - * a setting in the algorithm, not a model assumption,
  - * hierarchical data: use exchangeable type ($\sim$ equal correlations $\rightarrow$ next slide),
- invented in the 1980s and very much used since.

[16] Recall that generalized linear models (GLMs) constitute a general class of models including logistic and Poisson regression, specified by a distribution (family) and a link function.
[17] "Marginal" refers to the mean across the population of clusters (herds).

Working Correlation Matrices

For a series $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ of observations on a subject, the correlation matrix $\mathrm{Corr}(\boldsymbol{Y})$ is the $n \times n$-matrix of all pairs of correlations ~ $\boldsymbol{Y}$ as a multivariate outcome.

Examples of working correlation matrices[18] commonly used with GEE (shown for 4 observations per subject):

- independence:
  ~ independent or uncorrelated obs.,
  or no assumptions about $\mathrm{Corr}(\boldsymbol{Y})$

$$\mathrm{Corr}(Y_i, Y_j) = 0 \quad \text{for } i \neq j,$$

- exchangeable:
  ~ hierarchical data structure,
  with $\rho = \mathrm{ICC}$ (also compound symmetry)

$$\mathrm{Corr}(\boldsymbol{Y}) = \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho & \rho & 1 & \\ \rho & \rho & \rho & 1 \end{pmatrix},$$

- autoregressive `ar(1)`:
  ~ first order autoregressive,
  for repeated measures, has
  decaying correlation with time

$$\mathrm{Corr}(\boldsymbol{Y}) = \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho^2 & \rho & 1 & \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix},$$

- Toeplitz:
  ~ "stationary" (ρ depends on "distance" only)

$$\mathrm{Corr}(\boldsymbol{Y}) = \begin{pmatrix} 1 & & & \\ \rho_1 & 1 & & \\ \rho_2 & \rho_1 & 1 & \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix}.$$

[18] As the matrices are symmetric, for clarity the values above the diagonal are left blank.

GEE Inference and Variants

GEE settings:

- strongly recommended to use robust standard errors instead of model-based ones (caution: not the Stata default),
- note: GEE with independent correlation matrix $\sim$ robust variance estimation!,
- hypothesis testing: Wald tests mostly used (other tests exist),
- parameter estimates and SEs are asymptotically unbiased (when # clusters is large) under weak conditions,[19]
- performance in finite samples: standard guideline on sample size is that at least 30 clusters is needed to avoid biases,

Different versions of GEE for GLMs:

- differ by algorithms and settings for the part of the algorithm dealing with correlation structure – one important example: alternating logistic regression (ALR),
  - GEE-type procedure based on odds-ratios rather than working correlations (arguably more appropriate for binary data),
  - allows for either 2 or 3 hierarchical levels,
  - implemented in SAS and R, but not Stata.

[19] Importantly, misspecification of the working correlation matrix does not invalidate the estimates, but may lead to some loss of efficiency.

Choice of Correlation "Model"

How to choose the "best" (working) correlation structure for GEE?

- in principle[20], the QIC statistic (an analog of the AIC) can be compared between different correlation structures, and the one with smaller QIC value is preferable,
- guidelines (from Hardin and Hilbe (2003), *Generalized Estimating Equations*, 1st ed., p. 141 − 42):
  - for small cluster size and complete data, use unstructured,
  - for repeated measures over time, use a structure with time dependence, e.g. `ar(1)`,
  - for 2-level hierarchical structure, use exchangeable,
  - for small number of clusters, independence structure may be best,
  - if more than one structure meets the guidelines, use the QIC statistic to choose between them,
  - unstructured correlations may be used for exploratory analysis.
- another recommendation (Hardin & Hilbe): alternating logistic regression is preferable to ordinary GEE for logistic regression if "the focus of the analysis does include the association parameters" (i.e., association within clusters).

[20] My (Henrik's) experience with the QIC statistic has been mixed, and I don't feel comfortable with using QIC to choose the correlation structure.

Summary for GEE

Advantages and disadvantages of (classical) GEE method:

+ no assumptions about herd effects,
+ good/robust theoretical properties (with robust SE),
+ computationally feasible for large data sets,
+ possible/necessary to use knowledge of clustering,

+/− population-averaged instead of cluster-specific estimates,[21]

− less flexible with respect to multiple levels,
− no direct modelling of correlation structure.
− statistical choice of "working correlation structure" less clear,
− choice between different GEE variants remains "subjective".

Results (exchangeable correlation structure):
very close to linear mixed model (continuous data), but population-averaged estimates for binary data: 0.559 (.177) and 0.569 (.042) for herd- and cow-level $\boldsymbol{X}$.

[21] With identity link, e.g. linear models, there is no distinction between PA and CS estimates
⇒ GEE is perfectly valid for normally distributed outcomes, but often linear mixed models are preferred, due to their more clear-cut inference and larger flexibility.

Summary of Analyses for Simulated Datasets

Collected results for the 4 datasets (two outcomes, two predictor levels):

Estimate (SE) / Model/approach	Continuous outcome (true = 5) herd-level X	Continuous outcome (true = 5) cow-level X	Binary outcome (true = 0.693)[a] herd-level X	Binary outcome (true = 0.693)[a] cow-level X
unadjusted	3.557 (.200)	4.982 (.199)	0.529 (.042)	0.586 (.042)
mixed	3.796 (1.50)	4.968 (.149)	0.620 (.204)	0.697 (.046)
fixed effects	n/a	4.968 (.149)	n/a	0.704 (.046)
stratification	n/a	not discussed	n/a	0.698 (.046)
robust SE	3.557 (1.71)	4.982 (.142)	0.529 (.211)	0.586 (.044)
GEE (exch. corr)	3.797 (1.49)	4.968 (.141)	0.559 (.177)	0.569 (.042)

[a] cluster-specific true value, population-averaged equivalent ≈ 0.597 (computed as $0.693/\sqrt{1 + 0.346 \cdot 1} = 0.597$)

Extra notes:

- cow-level predictor: due to large sample size for X, all estimates are close to their (respective) true values,
- herd-level predictor: with an effective sample size of only 50 replicates per group, no close agreement with the true values can be expected.

Summary of Analyses for Pig Data

Estimates and SE (on logit scale):

Model	effect of ar_g1 Coef.	effect of ar_g1 SE	intercept Coef.	intercept SE
ordinary LR	0.647	0.220	-0.145	0.156
fixed effects LR	0.365	0.268	n/a	
Mantel-Haenszel	0.346	0.261	n/a	
robust variance	0.647	0.267	-0.145	0.279
random effects LR	0.437	0.258	0.020	0.301
GEE (exch corr.)	0.354	0.215	0.018	0.271

Conclusion:

- estimate by random effects model (GLMM) somewhat larger than for GEE; more than explained by being a cluster-specific estimate, as seen from (10bL−7)

$$0.437/\sqrt{1 + 0.346 \cdot 0.877} = 0.383,$$

  but not critically off considering the SE,
- Mantel-Haenszel and fixed effects estimates should be closer to GLMM than GEE estimates; some herd-level confounding (for ar_g1) appears to exist in the data,
- ordinary LR and robust SE considerably off, probably due to herd confounding.

MEANS/MARGINS FOR (LINEAR) MIXED MODELS

Basic fact: decisions are required (or implicit) on how to deal with the random effects.

Simplest situation: predictions on same scale as the random effects (say $\boldsymbol{u}$):

- setting $\boldsymbol{u}=0$ corresponds to predictions that are means in the random effects distribution,[22]
  - similar to setting $\boldsymbol{\varepsilon} = 0$ for prediction in linear models,
  - assumes prediction is for new "cluster" from population, as opposed to cluster(s) in the dataset,[23]
  - gives largest uncertainty (SE) for predictions because nothing is known about random effects from the data,
  - usual software default, e.g. in Stata's `margins` command.

More complex situation: (non-linear) transformation of means, e.g. if outcome was transformed for mixed model analysis:

- similar to linear models, where (back)transforming means gives medians, not means,
- for $\boldsymbol{u}=0$, interpret transformed values as medians,
- exact means can be obtained analytically (using formulae) for some transformations and generally by simulation; in practice, the median interpretation is often sufficient and satisfactory.

[22] With the usual assumption that random effects have mean 0, e.g. $\boldsymbol{u} \sim \mathrm{N}(0, \sigma_h^2)$.
[23] Sometimes termed referred to as "broad" and "narrow" inference spaces, respectively; Littell et al. (2006), *SAS for Mixed Models*, 2nd ed., SAS Publishing.

Means/Margins for GLMMs and GEE

Main issue: distinction between CS (cluster-specific) and PA (population-averaged) interpretations,

- PA predictions usually desired for broad[23] inference space,
  - directly with GEE (e.g. Stata's `margins` command),
  - caution needed (see below) for GLMMs,
- CS predictions may be of interest for specific clusters (included in the data), e.g. farms or schools — only available for GLMMs (technical: VER2/MER Ex. 22.10).

Common software limitations (Stata version $\leq$ 13, SAS, R): no predictions on original scale in GLMMs,[24]

- can work around these by manual backtransformation, but it will not give proper PA predictions: medians instead of means (as discussed on previous page), and thus not truly population-averaged,
- analytical adjustment to achieve PA estimates may be possible, e.g. in logistic regression using the formula on slide 10bL−7.

[24] In Stata 14+, only the `meqrlogit` and `meqrpoisson` commands do not allow estimation on original scale.