

AS14

GLMM and GEE for Categorical Repeated Measures – Linking SAS and R

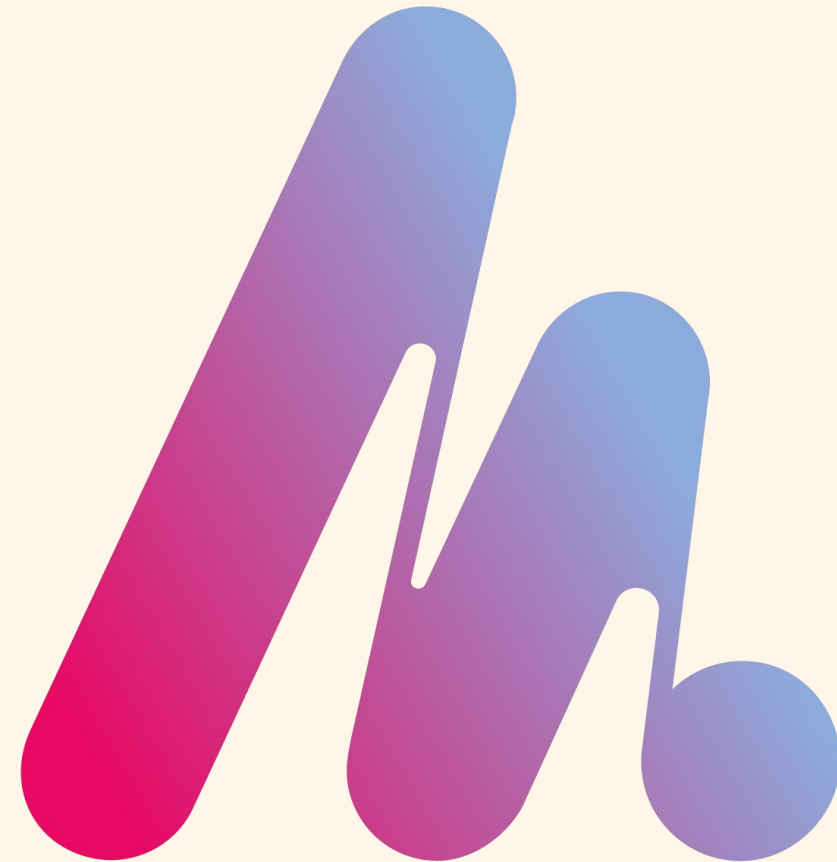
PHUSE Connect EU 2025

Miriam Amor Miriam.amor@veramed.co.uk

VeraMed

Agenda

- 01** Introduction
- 02** **GEE & GLMM**
An Overview
- 03** **GEE**
Binary outcome
- 04** **GLMM: GHQ & Laplace**
Binary outcome
- 05** **GEE & GLMM**
Outcome > 2 categories
- 06** **Conclusions**
- 07** **References**





Introduction

Introduction

Insert subhead for slide – 1 line only



R has gained significant traction and is often routinely used alongside **SAS**.



As a result, programmers are expected to understand potential **discrepancies in statistical outputs across these tools**, especially in statistical models.



To support this transition, the PHUSE **CAMIS** [1] provides a repository that documents and compares statistical methodologies across software.



This presentation explores how **GLMM** and **GEE** models are programmed in both SAS and R.



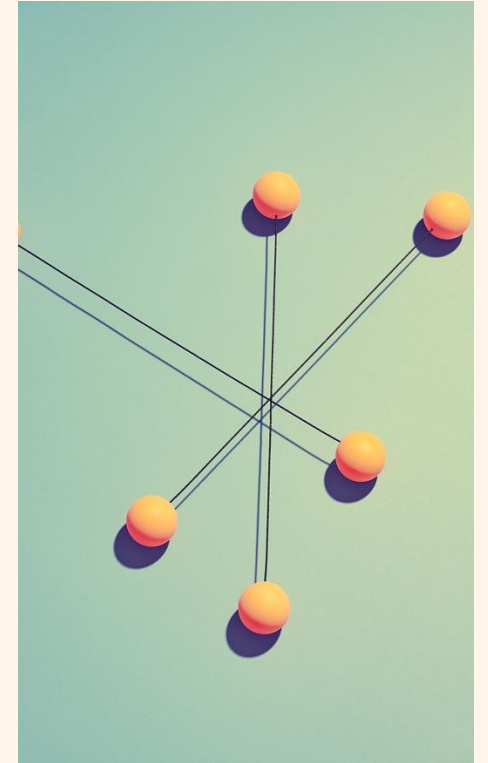
GEE & GLMM

An Overview

GEE & GLMM

An Overview

- Both Generalized Linear Mixed Model (**GLMM**) and Generalized Estimating Equations (**GEE**) methods extend the Generalized Linear Model (GLM) framework using **link functions** that relate the predictors to transformed outcome variable.
- Link functions:
 - Dichotomous response variables: **probit**
 - In case of rare events **complementary log-log** may be preferable.
 - Outcomes with more than two categories:
 - Ordinal: **cumulative logit**
 - Nominal: **generalized logit**



GEE & GLMM

An Overview



	GEE	GLMM
Model Type	Marginal (Population-averaged effects)	Conditional (Subject-average effects)
Use case	Population-level inference	Individual-level inference
Random Effects	Not explicitly included	Explicitly Included
Correlation	Working correlation structure	Modelled via random effects
Estimation method	Quasi-likelihood via iterative solving of estimating equations	Maximum Likelihood, based on approximations. - Gauss Hermite Quadrature (GHQ) - Laplace Pseudo-likelihood*

*Linear approximations instead of likelihood, making it less accurate for binary outcomes.

GEE & GLMM

An Overview: SAS procedures and R functions



Model	Software	Procedure/ Function	SE	Correlation Matrix	DF	Link Function			
						Probit	Logit	Clogit	Glogit
GLMM	SAS	PROC GLIMMIX	Model-Based	NA	B-W	✓	✓	✓	✓
	R	Lme4::glmer	Model-Based	NA	Inf.	✓	✓	✓	✓
		glmmTMB::glmmTMB	Model-Based	NA	Inf.	✓	✓	✗	✗
		GLMMadaptive::mixed_model	Model-Based	NA	Inf.	✓	✓	✗	✗
GEE	SAS	PROC GENMOD	Sandwich	Independent	NA	✓	✓	✓	✗
		PROC GEE	Sandwich	Independent	NA	✓	✓	✓	✓
	R	gee::gee	Sandwich	Independent	NA	✓	✓	✗	✗
		geepack::geeglm	Sandwich	Independent	NA	✓	✓	✗	✗
		ordLORgee	Sandwich	Exchangeable	NA	✗	✗	✓	✗
		nomLORgee	Sandwich	Exchangeable	NA	✗	✗	✗	✓

B-W: Between-Within. Inf: infinite. Sandwich Robust S.E. Estimator.

Model-Based is the SE estimated by the model (also called naive in R).



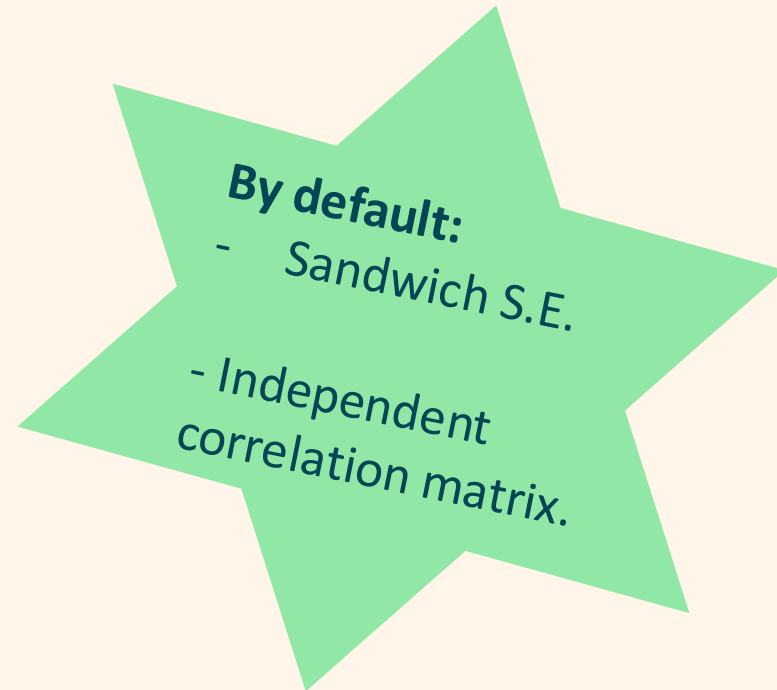
GEE

Binary outcome

GEE

Binary outcome

- **SAS:**
 - **PROC GEE or PROC GENMOD** (similar syntax).
 - Model-based SE with *modelse* option
- **R:**
 - **geepack::glm:**
 - requires outcome to be numeric (0,1)
 - **gee::gee**
 - It also displays model-based SE.
- **Results:**
 - Similar results in estimates and SE (Minor differences, attributable to differences in rounding.)



GEE: Binary outcome

Binary outcome

SAS	<pre>proc gee/proc genmod data=resp; class trt(ref="P") avisitn(ref='1') usubjid; model outcome(event='1') = trt avisitn trt*avisitn/ dist=bin link=logit; lsmeans trt*avisitn/cl exp ilink oddsratio diff; repeated subject=usubjid/corr=ind; run;</pre>
R	<pre>#Create numeric outcome (values: 1/0). resp\$outcome_num<-as.numeric(resp\$outcome)-1 library(geepack) model <- geeglm(outcome_num ~ trt + avisitn +trt*avisitn, id=usubjid, data=resp, family=binomial(link='logit'), corstr="independence") library(gee) model<- gee(outcome ~ trt + avisitn + trt*avisitn, id=usubjid, data=resp, family=binomial(link='logit'), corstr = "independence")</pre>

Prob. & OR:
SAS: LSMEANS option
R: emmeans function

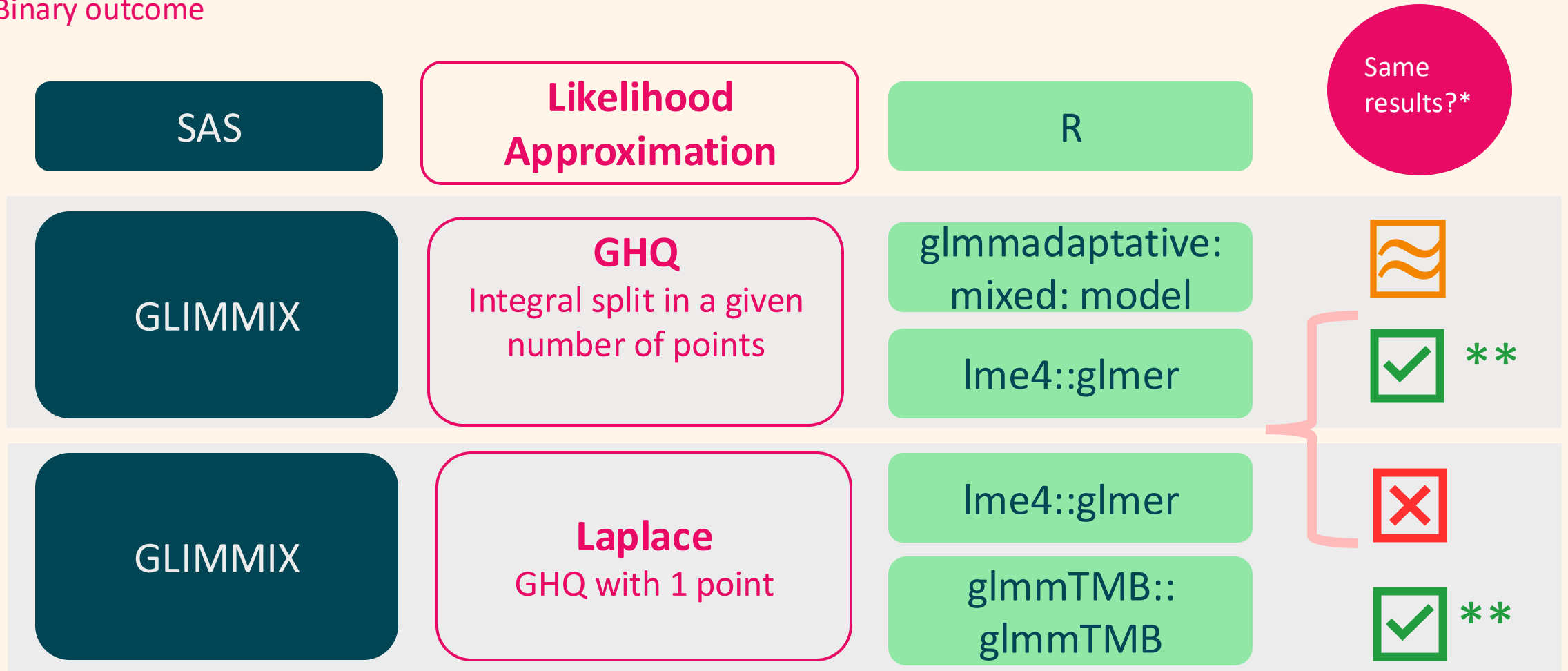
Source data: [SAS Help Center](#) [2].

GLMM: GHQ & Laplace

Binary outcome

GLMM: GHQ & Laplace

Binary outcome



*Using example data: [SAS Help Center](#).

** After aligning options between SAS and R.

GLMM: GHQ & Laplace

Binary outcome

GHQ	SAS	<pre>proc glimmix data=resp method=quad(qpoints=5); class trt(ref="A") avisitn(ref='1') usubjid; model outcome=trt avisitn trt*avisitn/dist=bin link=logit solution ddfm=none; random intercept /subject=usubjid; run;</pre>
	R	<pre>model <- glmer(outcome ~ trt + avisitn + trt*avisitn + (1 usubjid), data = resp, family = binomial(link = "logit"), nAGQ=5)</pre>
Laplace	SAS	<pre>proc glimmix data=resp method=Laplace; /*Or quad(qpoints=1)*/ class trt(ref="A") avisitn(ref='1') usubjid; model outcome=trt avisitn trt*avisitn/dist=bin link=logit solution ddfm=none; random intercept /subject=usubjid; run;</pre>
	R	<pre>model<- glmmTMB(outcome ~ trt + avisitn + trt*avisitn + (1 usubjid), data = resp, family = binomial(link = "logit"))</pre>

To match R ddf.
(SAS: B-W by default)

GLMM: GHQ & Laplace

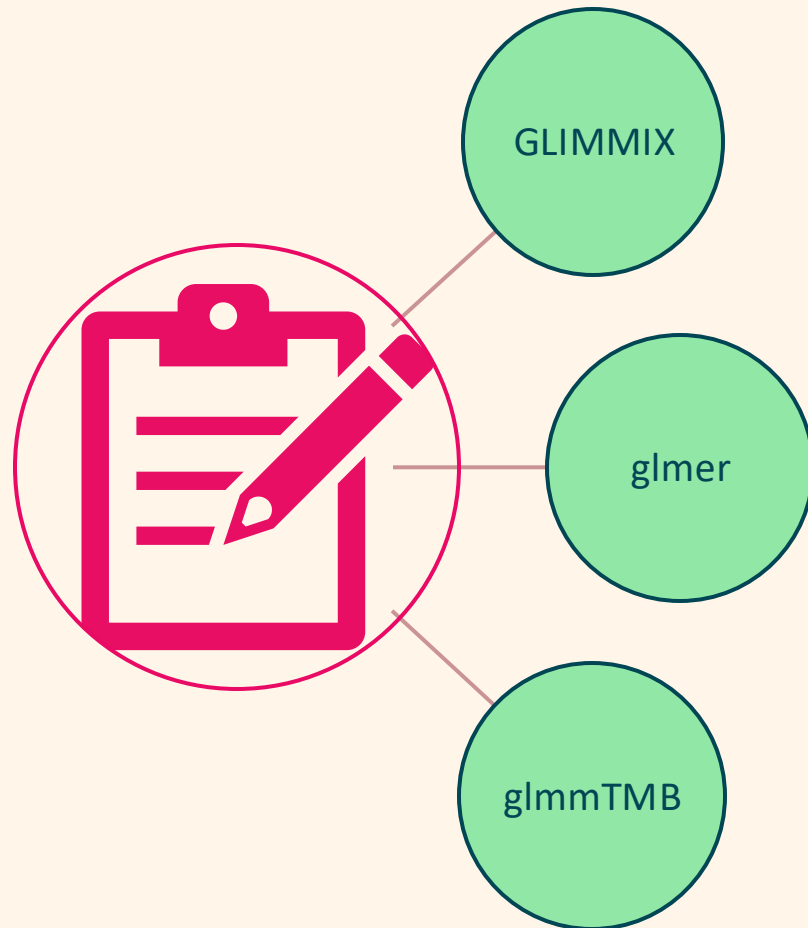
Binary outcome



- Previous results match with **model-based (naïve) S.E.** and **infinite ddff**. However...
 - Li, P., & Redden, D. T. (2015) [3], suggested that the **Between-Within** denominator degrees of freedom approximation method when using GLMMs in cluster randomized trials with binary outcomes and a small number of heterogeneous clusters.
 - Additionally, the FDA advise: *“sponsors to consider using of **robust standard error method such as the Huber-White “sandwich” standard error**, particularly when the model does not include treatment by covariate interaction”* [4].

GLMM: GHQ & Laplace

Binary outcome



Options:

- Empirical
- ddfm=betwithin (by default)

Based on Li and Redden 2015 [3]

Additional functions

- merDeriv::sandwich
- Parameter::dof_betwithin

Implements a heuristic and does not return exactly same results as in based on Li and Redden 2015 [3]

Additional functions

- clubSandwich::vcovHC
- parameter::dof_betwithin

GLMM: GHQ & Laplace

Binary outcome



SAS

```
proc glimmix data=resp method=quad(qpoints=5) empirical;  
  class trt(ref="P") avisitn(ref='1') usubjid;  
  model outcome=trt avisitn trt*avisitn/dist=bin link=logit solution ddfm=betwithin;  
  random intercept /subject=usubjid;  
run;
```

R

```
model <- glmer(outcome ~ trt + avisitn + trt*avisitn + (1 | usubjid),  
              data = resp,  
              family = binomial(link = "logit"),  
              nAGQ=5)  
  
#Get parameter estimation  
est<-fixef(model)  
#Get Sandwich covariance matrix  
vcov<-merDeriv::sandwich(model)  
  
#Get S.E. (the diagonal from the covariance matrix)  
se_sw0<- sqrt(diag(sandwich(model)))  
se_sw <-head(se_sw0, -1)  
  
#Re-calculate p-values using different Between-Within df  
t <- est/se_sw  
pvalue <-2*pt(q=abs(t), df=dof_betwithin(model), lower.tail=FALSE)
```

Heuristic: does not return exact results.

GLMM: GHQ & Laplace

Binary outcome



	SAS								R					
DF: Inf	Effect	trt	avisitn	Estimate	StdErr	DF	tValue	Probt						
	Intercept		—	-0.0617	.5345	I	-0.1154	0.9081	(Intercept)	-0.0617	0.5378	Inf	-0.1146	0.9087
	trt	A	—	1.5750	.7868	I	2.0019	0.0453	trtA	1.5750	0.7914	Inf	1.9901	0.0466
	avisitn		2	-0.8719	.5391	I	-1.6173	0.1058	avisitn2	-0.8720	0.5392	Inf	-1.6173	0.1058
	avisitn		3	-0.2873	.6102	I	-0.4708	0.6378	avisitn3	-0.2873	0.6102	Inf	-0.4708	0.6378
	avisitn		4	-0.4317	.5164	I	-0.8361	0.4031	avisitn4	-0.4318	0.5164	Inf	-0.8362	0.4031
	trt*avisitn	A	2	1.0438	.8219	I	1.2700	0.2041	trtA:avisitn2	1.0440	0.8220	Inf	1.2700	0.2041
	trt*avisitn	A	3	0.6361	.8588	I	0.7406	0.4589	trtA:avisitn3	0.6362	0.8589	Inf	0.7407	0.4589
	trt*avisitn	A	4	-0.2200	.7939	I	-0.2771	0.7817	trtA:avisitn4	-0.2200	0.7939	Inf	-0.2771	0.7817
DF: Between-Within	Effect	trt	avisitn	Estimate	StdErr	DF	tValue	Probt						
	Intercept		—	-0.0617	.5345	109	-0.1154	0.9083	(Intercept)	-0.0617	0.5378	324	-0.1146	0.9088
	trt	A	—	1.5750	.7868	109	2.0019	0.0478	trtA	1.5750	0.7914	324	1.9901	0.0474
	avisitn		2	-0.8719	.5391	327	-1.6173	0.1068	avisitn2	-0.8720	0.5392	324	-1.6173	0.1068
	avisitn		3	-0.2873	.6102	327	-0.4708	0.6381	avisitn3	-0.2873	0.6102	324	-0.4708	0.6381
	avisitn		4	-0.4317	.5164	327	-0.8361	0.4037	avisitn4	-0.4318	0.5164	324	-0.8362	0.4037
	trt*avisitn	A	2	1.0438	.8219	327	1.2700	0.2050	trtA:avisitn2	1.0440	0.8220	324	1.2700	0.2050
	trt*avisitn	A	3	0.6361	.8588	327	0.7406	0.4594	trtA:avisitn3	0.6362	0.8589	324	0.7407	0.4594
	trt*avisitn	A	4	-0.2200	.7939	327	-0.2771	0.7819	trtA:avisitn4	-0.2200	0.7939	324	-0.2771	0.7819

Small differences in S.E

Differences in ddf.
dof_betwithin
↓
based on approximation.



GEE & GLMM

Outcome with more than 2 categories

GEE & GLMM

Outcome with more than 2 categories

- The **multinomial** distribution is used, and link function:
 - **cumulative** to ordinal data.
 - **logit** to fit to nominal data.
- In **SAS**, similar syntax used for GEE and GLMM models can be applied.
- In **R**:
 - Functions described before do not support multinomial outcomes.
 - Two different functions are available in the **multgee** package for **GEE** models.
 - ordLORgee for ordinal
 - nomLORgee for nominal variables.
- Equivalent functions for **GLMM** have not been identified in a frequentist framework

Correlation matrix by default: Exchangeable

GEE & GLMM

Outcome with more than two categories

SAS	Ordinal	<pre>proc gee data=resp ; class trt(ref="A") avisitn(ref='1') usubjid; model respord=trt avisitn trt*avisitn/ repeated subject=usubjid /corr=ind; run;</pre>	
	Nominal	<pre>proc gee data=resp ; class trt(ref="A") avisitn(ref='1') usubjid; model respnom(event='Liver')=trt avisitn trt*avisitn/ repeated subject=usubjid /corr=ind; run;</pre>	
R	Ordinal	<pre>model <- ordLORgee(formula = respord ~ trt + avisitn + trt*avisitn, data = resp, id = usubjid, repeated = avisitn, LORstr = "independence")</pre>	Correlation structure set to 'Independence'.
	Nominal	<pre>model <- nomLORgee(formula = respnom ~ trt + avisitn + trt*avisitn, data = resp, id = usubjid, repeated = avisitn, LORstr = "independence")</pre>	



Conclusions

From the specific to
the general

Conclusions



GEE

- Correlation matrix
- = results (Sandwich SE)
- Outcome > 2 categories



GLMM

- Random effects
- GHQ, Laplace approx.
- = results (Naïve SE, Inf. DF)
- $\approx$ results (Sandwich SE, B-W DF)
- Outcomes > 2 categories only in SAS



OR

- LSMEANS in SAS
- emmeans in R



SAS vs R

- Default settings
- Computational methods
- SAS options/ R additional functions
- [CAMIS repository](#) [1] documents differences between software.



References

References



- [1] [Comparing Analysis Method Implementations in Software \(CAMIS\)](#)
- [2] [SAS Institute Inc.. SAS Help Center. The GEE procedure.](#)
- [3] [Li, P., & Redden, D. T. \(2015\). Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. BMC Medical Research Methodology, 15, 38.](#)
- [4] [U.S. Food and Drug Administration. \(2023\). Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products: Guidance for Industry. Center for Drug Evaluation and Research \(CDER\), Center for Biologics Evaluation and Research \(CBER\).](#)
- [3] [SAS/STAT® 13.1 User's Guide The GENMOD Procedure.](#)
- [5] [SAS/STAT® 13.1 User's Guide The GLIMMIX Procedure](#)
- [6] [Stack Overflow \[Internet\]. 2008 \[Last visited: 2025 Sep 30\]](#)
- [7] [Brooks, M. E., et al. \(2025\). glmmTMB: Generalized Linear Mixed Models using Template Model Builder \(Version 1.1.12\) \[R package manual\]. The Comprehensive R Archive Network \(CRAN\). <https://cran.r-project.org/web/packages/glmmTMB/glmmTMB.pdf>](#)
- [8] [Documentation of package parameters. dof_betwithin](#)
- [9] [Touloumis A. \(2015\). "R Package multgee: A Generalized Estimating Equations Solver for Multinomial Responses." Journal of Statistical Software.](#)



Any Questions?