

Index of Lecture 11–L

Page	Title
1	GEE cautions and developments
2	Working correlation structure
3	Alternating logistic regression (ALR)
4	ALR vs. classical GEE
5	ALR and GEE for cross-over trials
6	Example: Indonesian children study
7	Models for Indonesian children study (Example 8.7 in Diggle et al., 2002)

— in addition to overheads from Davis (2002):
642,647–49,651–53,664–65,668–72,698
and 600,607–09 from 10–L

GEE CAUTIONS AND DEVELOPMENTS

Notes on classical GEE:

- hypothesis tests,
- performance (in finite samples),
- cautions, in particular time-varying covariates,
- choice of working correlation matrix structure (next slide),
- diagnostics and goodness-of-fit tests (logistic regression): proposed in the literature but not implemented yet,
- sample size calculations (Liu & Liang, 1997): general theoretical framework but only new example is a 2×2 cross-over design with binary response.

Other (newer) versions of GEE

- substitution of Pearson residuals by other residuals,
- GEE2 — different estimation method for correlations (using a second set of estimating equations),
- alternating logistic regression (ALR) — alternative form of GEE, but only for logistic and ordinal regression.

GEE for ordinal categorical responses:

- only in SAS (to my knowledge), and only with independence working correlation structure.

WORKING CORRELATION STRUCTURE

Tools to aid choice of correlation structure:

(none of them implemented as standard in software)

- two statistics c_1 and c_2 indicate closeness of working correlation matrix to true correlation matrix — not in common use,
- modified Akaike criterion for GEE (QIC) — promising approach and does not seem so hard to implement,
- variogram or “lorelogram” as diagnostic tools (two different papers).

Guidelines (Hardin & Hilbe, 2003):

- if the cluster size is small and the data are complete, use unstructured,
- if repeated measures over time, use a structure with time dependence,
- for hierarchical structure, use exchangeable,
- if the number of clusters is small, the independence model may be best (but use the robust/sandwich variance estimate),
- if more than one structure meets the guidelines, use the QIC to choose between them.

ALTERNATING LOGISTIC REGRESSION (ALR)

To measure association between two binary outcomes:

- we often use the odds-ratio, $\psi = p_{11}p_{00}/p_{01}p_{10}$, where $p_{ij} = P(Y_1 = i, Y_2 = j)$ for $i, j = 0, 1$;
- we rarely use the correlation $\rho = \rho(Y_1, Y_2)$, because linear association makes more sense for continuous than binary variables.

Alternating logistic regression:

- assumes same model for $\text{logit}(p)$ as classical GEE,
- specifies the association between observations within the same cluster in terms of the log odds-ratio ($\ln(\psi)$) instead of the correlation,
- allows to specify “working models” for the log odds-ratio, similarly to the working correlation matrices:
 - * 2-level hierarchy: exchangeable type where all odds-ratios within a cluster are the same,
 - * repeated measures: unstructured also possible,
 - * 3-level hierarchy: exchangeable (two correlations),
 - * may also depend on predictors (examples hereof are rare. . .).

ALR vs. CLASSICAL GEE

Recommendations:

- ALR gives a better estimate of the association within a cluster than ordinary GEE and an associated SE
⇒ generally recommended (Hardin & Hilbe, 2003),
- small difference in inference about fixed effects between approaches based on correlation and odds-ratio (Davis, 2002),
- only ALR is discussed in detail for logistic regression (Diggle et al., 2002).

Software for ALR and GEE:

- ALR: only SAS and S-Plus/R (**a1r** library), seems to give same results.
- GEE: Stata, SAS and S-Plus/R (**gee** library), gives different results for unbalanced data with time-dependent correlation structures:
 - * Stata requires **force** option,
 - * SAS, Stata and R give slightly different results, the implementation in R seems less reliable to me.

ALR AND GEE FOR CROSS-OVER TRIALS

2x2 trial:

- fixed effects (intercept, tx, period, interaction) estimates and SE identical on two decimals,
- estimates of clustering (without interaction):
working correlation : 0.64, log odds-ratio : 3.56 (.81).

3x3 trial: ALR/GEE logistic regression estimates:

Parameter		ALR		GEE	
		exch.	unstr.	exch.	unstr.
intercept		-1.08 (.32)	-1.10 (.32)	-1.08 (.32)	-1.10 (.32)
β (period)	2	0.42 (.41)	0.38 (.41)	0.42 (.41)	0.38 (.41)
β (period)	3	0.59 (.46)	0.55 (.45)	0.59 (.46)	0.55 (.45)
β (tx)	B	2.09 (.42)	2.09 (.42)	2.10 (.42)	2.10 (.42)
β (tx)	C	2.07 (.42)	2.12 (.42)	2.07 (.42)	2.13 (.42)
β (prev. tx)	B	-0.15 (.51)	-0.09 (.50)	-0.14 (.51)	-0.09 (.50)
β (prev. tx)	C	-0.92 (.44)	-0.86 (.43)	-0.93 (.44)	-0.87 (.43)
clustering:	(1,2)	-0.22 (.38)	-0.96 (.59)	-0.03	-0.17
log odds-ratio/	(1,3)	-0.22 (.38)	0.11 (.70)	-0.03	0.03
working corr.	(1,2)	-0.22 (.38)	0.29 (.69)	-0.03	0.04

- almost perfect agreement ALR/GEE for fixed effects,
- similar pattern in ALR/GEE cluster parameters,
- small difference between exchangeable and unstructured:
 - * fixed effects very close,
 - * one notable, negative correlation (times 1 and 2), others close to zero.

EXAMPLE: INDONESIAN CHILDREN STUDY

Cohort study involving 275 children (subset of full dataset):

- examined through 6 consecutive quarters,
- outcome is respiratory infection (no/yes),
- predictors of primary interest: age, Xerophthalmia (ocular manifestation of vitamin A deficiency),
- full list of predictors:

Variable	Description	Values
gender	gender	0/1
height	age-adjusted height	-23–25 (% of standard)
age	age in years	2–14
cos	seasonal cosine ^a	-1,0,1
sin	seasonal sine ^a	-1,0,1
xerop	Xerophthalmia	0/1
age	age centered at 3 years	-32–50 months
age0	age at study onset	-32–44 months
follow-up	time since study onset	0–15 months

^a period of one year, starting in winter

Descriptive statistics per visit:

Prevalences (%)	Visit (season; 1=summer)					
	1	2	3	4	5	6
Resp. infection	12.6	4.6	7.3	3.8	14.9	9.4
Xerophthalmia	3.9	6.1	6.2	5.5	3.1	3.0
No. of children	230	214	177	183	195	201

MODELS FOR INDONESIAN CHILDREN STUDY

(EXAMPLE 8.7 IN DIGGLE ET AL., 2002)

Model 1: data from first visit only,

- cross-sectional estimates only,
- no clustering/repeated measures to take into account.

Models 2–4: data from all visits,

- fit by ALR with exchangeable correlations,
- model 2: no separation of cross-sectional and longitudinal effects (age, Xerophthalmia),
- models 3–4: simultaneous fit of
 - * age at entry (cross-sectional effect),
 - * follow-up time (longitudinal effect),
 - highly correlated with seasonal effect (because only 6 seasons), and becomes non-significant when season is included (model 4).