

# Generalized Shared-parameter Models and Missingness at Random

An Creemers<sup>1</sup>   Niel Hens<sup>1,2</sup>   Marc Aerts<sup>1</sup>   Geert Molenberghs<sup>1,3</sup>

Geert Verbeke<sup>3,1</sup>   Michael G. Kenward<sup>4</sup>

<sup>1</sup> *I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium*

<sup>2</sup> *Centre for Health Economics Research and Modeling Infectious Diseases (CHERMID),*

*Centre for the Evaluation of Vaccination (WHO Collaborating Centre),*

*Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium*

<sup>3</sup> *I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

<sup>4</sup> *Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London WC1E7HT, UK*

## Supplementary Materials

### A SAS Implementation

Assume we have a dataset, 'toenailc02', with variables 'idnum' an indicator for subject, 'treat' a (binary) indicator for treatment ('0' for standard arm, '1' for the experimental arm), and 'time' time in months (0, 1, 2, 3, 6, 9, and 12 months). Further, the actual outcome, unaffected nail length, and the dropout indicators are stacked into a single variable 'respos', with 'resptype' indicating whether the outcome listed is of the former ('1') or the latter ('0') kind. It is sometimes convenient to dispose of character variables indicating the outcome distribution, 'dist', and the link function chosen 'link', respectively. The need for these two variables arises when one makes use of, for example, the SAS procedure GLIMMIX, but they are superfluous with NLMIXED, the procedure used here.

Data for three typical subjects are as follows:

```
idnum  treat  time  respos  resptype  dist  link
```

Generalized Shared-parameter Models and MAR

2	0	0	4	1	GAUS	IDEN
2	0	1	6	1	GAUS	IDEN
2	0	2	7	1	GAUS	IDEN
2	0	3	9	1	GAUS	IDEN
2	0	6	13	1	GAUS	IDEN
2	0	9	0	1	GAUS	IDEN
2	0	12	.	1	GAUS	IDEN
2	0	0	0	0	BINA	LOGI
2	0	1	0	0	BINA	LOGI
2	0	2	0	0	BINA	LOGI
2	0	3	0	0	BINA	LOGI
2	0	6	0	0	BINA	LOGI
2	0	9	0	0	BINA	LOGI
2	0	12	1	0	BINA	LOGI
3	0	0	0	1	GAUS	IDEN
3	0	1	1	1	GAUS	IDEN
3	0	2	2	1	GAUS	IDEN
3	0	3	3	1	GAUS	IDEN
3	0	6	9	1	GAUS	IDEN
3	0	9	4	1	GAUS	IDEN
3	0	12	2	1	GAUS	IDEN
3	0	0	0	0	BINA	LOGI
3	0	1	0	0	BINA	LOGI
3	0	2	0	0	BINA	LOGI
3	0	3	0	0	BINA	LOGI

3	0	6	0	0	BINA	LOGI
3	0	9	0	0	BINA	LOGI
3	0	12	0	0	BINA	LOGI
18	0	0	0	1	GAUS	IDEN
18	0	1	0	1	GAUS	IDEN
18	0	2	0	1	GAUS	IDEN
18	0	3	.	1	GAUS	IDEN
18	0	6	.	1	GAUS	IDEN
18	0	9	.	1	GAUS	IDEN
18	0	12	.	1	GAUS	IDEN
18	0	0	0	0	BINA	LOGI
18	0	1	0	0	BINA	LOGI
18	0	2	0	0	BINA	LOGI
18	0	3	1	0	BINA	LOGI

Subject #2 drops out at the last occasion, #3 completes the study, and #18 drops out at the third month.

Since all subjects are observed at the baseline occasion (month 0), it is best to remove the corresponding line from the dataset:

```
data help;
set m.toenailc02;
if (resptype=0 and time=0) then respons=.;
run;
```

The NL MIXED procedure can then be invoked for parameter estimation and for prediction of unobserved measurements:

```

proc nlmixed data=help qpoints=20 maxiter=500;
parms beta0=3 beta1=-0.25 beta2=0.5 beta3=0.05
      gamma0=-3.5 gamma01=1 gamma1=0 gamma2=0 gamma3=0
sigma=2.5 tau=2;
if resptype=1 then do;
  mean = beta0 + b + beta1*treat + beta2*time + beta3*treat*time;
  dens = -0.5*log(3.14159265358) - log(sigma)
        - 0.5*(respons-mean)**2/(sigma**2);
  ll = dens;
end;
else if resptype=0 then do;
  eta = gamma0 + b*gamma01 + gamma1*treat + gamma2*time + gamma3*treat*time;
  p = exp(eta)/(1+exp(eta));
  ll = respons*log(p) + (1-respons)*log(1-p);
end;
model respons ~ general(ll);
random b ~ normal(0,tau*tau) subject=idnum;
estimate 'tau^2' tau*tau;
estimate 'gamma01^2*tau^2' gamma01*gamma01*tau*tau;
estimate 'sigma^2' sigma*sigma;
predict (beta0 + b + beta1*treat + beta2*time + beta3*treat*time)
      out=m.toenailc2mnr;
predict (beta0 + beta1*treat + beta2*time + beta3*treat*time)
      out=m.toenailc2mar;
run;

```

The actual model consists, of course, of two parts. Using branching statements, a Gaussian likelihood contribution is calculated for unaffected nail length, whereas a Bernoulli contribution is calculated for the dropout indicators. Thanks to the 'general' likelihood feature in the MODEL statement, the user

has a very large amount of freedom in model specification. Additional ESTIMATE statements lead to the direct determination of parameter estimates and standard errors for the variance components.

Finally, the two PREDICT statements determine the prediction of the unobserved measurements given the observed ones, using the model as it is, based on (33), or using the MAR counterpart, as in (34). Note that this is relatively simple, given the conditional independence assumption, as discussed in Section 7.1. Given that these predictions range over the entire measurement sequence and not only over the missing ones, we need to bring together the observed measurements until the time of dropout, and then switch to the predictions thereafter. Using the following code, this is done for the original MNAR model and for the MAR counterparts. These two output datasets are then merged into a single one.

```
data m.toenailc2mnar;
set m.toenailc2mnar;
mnar=respons;
if resptype=0 then mnar=.;
if (respons=. and resptype=1) then mnar=pred;
run;
```

```
data m.toenailc2mar;
set m.toenailc2mar;
mar=respons;
if resptype=0 then mar=.;
if (respons=. and resptype=1) then mar=pred;
run;
```

```
data m.toenailc2m;
merge m.toenailc2mnar m.toenailc2mar;
if resptype=0 then delete;
keep idnum treat time respons mnar mar;
```

```
run;
```

The final dataset can now be used as a basis to construct, for example, Figure 4.