

Log-Gamma Modeling of HIV Medication Adherence Data Using SAS

Tor Neilands, Ph.D.

UCSF Center for AIDS Prevention Studies

CAPS Methods Core Presentation

May 14, 2010

SAS Implementation of the example HLP study log-gamma model for non-adherence

Some preliminaries:

- Everything you will see in this talk should be considered a work in progress. Your questions and especially your feedback and suggestions for improvements or alternative approaches are most welcome.
- HLP Log-Gamma Model for Non-Adherence:
 - We first re-created the non-adherence variable in SAS:
 - `glmadh = 101 - pctadh1 ;`
 - It is necessary in SAS for the outcome to have values greater than zero for gamma regression modeling; values of zero or less will be dropped from analyses.
 - We created dummy variables `bdisleep1`, `bdisleep2`, and `bdisleep3` to represent BDI category 1 vs. 0, 2 vs. 0, and 3 vs. 0, respectively.
 - We also created a binary outcome, `y_binary`, whose value is zero if non-adherence was zero, one if non-adherence was 1% to 100%, and missing if non-adherence was missing.

Reprising the Stata analysis using SAS PROC GLIMMIX

- First, we replicated the Stata analysis using PROC GLIMMIX
- You could also do this analysis using PROC GENMOD. We chose GLIMMIX to make this demonstration consistent with later demonstrations that make use of repeated measures data.

```
proc glimmix data = temp ;  
  title "GLIMMIX gamma analysis of HLP data" ;  
  model glmadh = bdisleep1 bdisleep2 bdisleep3 /  
    solution dist = gamma link = log ;  
  estimate 'Exp(bdisleep1)' bdisleep1 1 / exp cl ;  
  estimate 'Exp(bdisleep2)' bdisleep2 1 / exp cl ;  
  estimate 'Exp(bdisleep3)' bdisleep3 1 / exp cl ;  
run ;
```

Results from PROC GLIMMIX

Fit Statistics

-2 Log Likelihood	17445.47
AIC (smaller is better)	17455.47
AICC (smaller is better)	17455.49

Parameter Estimates

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	2.1423	0.04707	2840	45.51	<.0001
bdisleep1	0.1691	0.06131	2840	2.76	0.0058
bdisleep2	0.2321	0.08353	2840	2.78	0.0055
bdisleep3	0.4352	0.1079	2840	4.03	<.0001
Scale	2.0094	0.04397	.	.	.

Estimates

Label	Exponentiated Estimate	Exponentiated Lower	Exponentiated Upper
Exp(bdisleep1)	1.1842	1.0501	1.3355
Exp(bdisleep2)	1.2613	1.0707	1.4857
Exp(bdisleep3)	1.5453	1.2506	1.9096

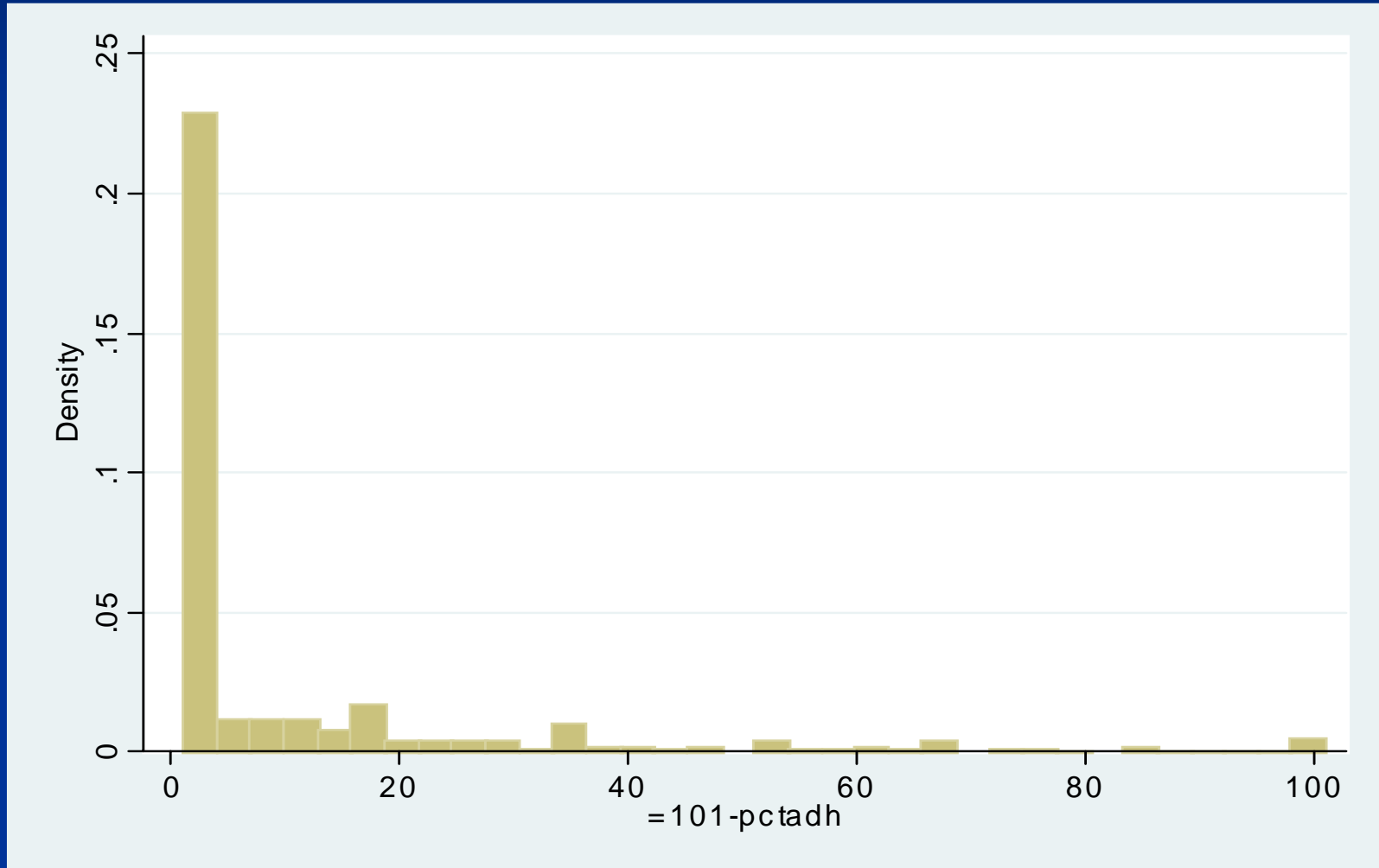
GLIMMIX Results

- Increasing steps in sleep disturbance are positively associated with non-adherence
- Point estimates are equivalent to those produced by `-glm-` in Stata
- Standard errors and confidence intervals are different. Equivalent SEs and CIs can be obtained by specifying the `robust` option in Stata and using the option `EMPIRICAL = CLASSICAL` in `PROC GLIMMIX`.
- If using `SAS PROC GENMOD` instead of `PROC GLIMMIX`, use a `REPEATED` statement with `SUBJECT` set equal to the subject ID variable.

Should we stop here?

- As one of my statistics teachers once said, “If you have one watch, you always know what time it is; if you have more than one watch, you never know what time it is”.
- We have already introduced the log-gamma model as an alternative to usual practice of logistic regression of a binary adherence outcome. Which of these two watches do we believe?
- What if we could combine the two approaches to yield a best-of-both-worlds scenario that represents a third approach to modeling HIV medication adherence data?
- At the risk of confusing ourselves at a higher level, we will consider that question next. Perhaps that can help us establish a plurality of accuracy (many wristwatches collectively may be even more accurate than one or two).

Distribution of Non-Adherence



Distribution of Non-Adherence

- The distribution of non-adherence appears to be *zero-inflated*. In other words, many participants were not non-adherent (i.e., they reported perfect adherence). 65% ($n = 1855$) reported perfect adherence in the HLP data.
- Could this zero inflation affect model results?
- To investigate this question, it would be helpful to have a method to jointly model the zero cases versus the non-zero cases using a logistic regression approach and simultaneously model the non-zero cases using a gamma distribution.

Zero-Inflated Log-Gamma Model

- This can be accomplished using the *zero-inflated log-gamma* (ZILG) model.
- How do we obtain ZILG model results? Unlike zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB), at the time of this presentation there is no canned software routine in SPSS, SAS, or Stata for estimating a ZILG.
- The analyst must write the likelihood function directly and use a general maximum likelihood estimation program such as Stata's `-ml-` or SAS's PROC NLMIXED to perform the estimation.

Log-Gamma Model in PROC GLIMMIX

- Start out by estimating a non-zero-inflated gamma analysis for just the non-zero cases using PROC GLIMMIX:

```
proc glimmix data = temp ;  
  title "GLIMMIX gamma analysis of HLP data for GLMADH > 0" ;  
  where glmadh > 1 ;  
  y = glmadh - 1 ;  
  model y = bdisleep1 bdisleep2 bdisleep3 /  
          solution dist = gamma link = log ;  
  estimate 'Exp(bdisleep1)' bdisleep1 1 / exp cl ;  
  estimate 'Exp(bdisleep2)' bdisleep2 1 / exp cl ;  
  estimate 'Exp(bdisleep3)' bdisleep3 1 / exp cl ;  
run ;
```

Log Gamma Model Results

```

-2 Log Likelihood      8366.95
AIC (smaller is better) 8376.95
AICC (smaller is better) 8377.01
    
```

Parameter Estimates

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	3.2108	0.05292	987	60.67	<.0001
bdisleep1	0.008780	0.06655	987	0.13	0.8951
bdisleep2	0.1003	0.08907	987	1.13	0.2604
bdisleep3	0.1164	0.1053	987	1.11	0.2690
Scale	0.7701	0.03114	.	.	.

Label	Exponentiated Estimate	Exponentiated Lower	Exponentiated Upper
Exp(bdisleep1)	1.0088	0.8853	1.1495
Exp(bdisleep2)	1.1055	0.9282	1.3167
Exp(bdisleep3)	1.1235	0.9138	1.3813

Logistic Model

- None of the BDI sleep category variables are statistically significant for participants with one or more percentage point of non-adherence.
- What about the odds of being in the perfect adherence group (i.e., zero non-adherence)?
- Use ordinary logistic regression to address this question

```
proc logistic data = temp ;  
  title "Logistics analysis of HLP data for GLMADH 0 vs. 1" ;  
  model y_binary(event='0') = bdisleep1 bdisleep2 bdisleep3 ;  
run ;
```

Logistic Results

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.8321	0.0722	132.6797	<.0001
bdisleep1	1	-0.2710	0.0924	8.5996	0.0034
bdisleep2	1	-0.2369	0.1247	3.6077	0.0575
bdisleep3	1	-0.5772	0.1559	13.7093	0.0002

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
bdisleep1	0.763	0.636	0.914
bdisleep2	0.789	0.618	1.008
bdisleep3	0.561	0.414	0.762

Summary of HLP Example Results

- Treating the outcome variable as non-adherence measured on a 1-101 percentage point scale yielded significant positive associations between each level of the BDI sleep variable and non-adherence, using a log-gamma generalized linear model.
- Splitting the outcome variable into perfect adherence (0) vs. some degree of non-adherence (1) using a logistic regression approach and then further modeling the amount of non-adherence in the non-adherent cases using a log-gamma generalized linear model suggests that poor sleep quality is associated with a lower odds of being perfectly adherent, but sleep quality is not significantly related to the amount of non-adherence among non-adherent cases.
- What about examining these effects jointly the same model?

Zero-Inflated Gamma Model in PROC NLMIXED

```
proc nlmixed data = temp qpoints = 15 lognote = 3 ;  
  parms a0=0.1 a1=0.1 a2=0.1 a3=0.1 b0=0.4 b1=0.2 b2=0.4 b3=0.4 log_theta = 0;  
  title "NLMIXED - Log-Gamma inflated model" ;  
  title2 "Syntax is based on Dale McLerran's posts on SAS-L" ;  
  y = glmadh-1;  
  
  linfp = a0 + a1*bdisleep1 + a2*bdisleep2 + a3*bdisleep3 ;  
  infprob = exp(linfp)/(1+exp(linfp)) ;  
  
  linp = b0 + b1*bdisleep1 + b2*bdisleep2 + b3*bdisleep3 ;  
  mu = exp(linp);  
  theta = exp(log_theta) ;  
  r = mu/theta ;  
  
  if y = 0 then ll = log(infprob) ;  
  else ll = log(1-infprob) - lgamma(theta) + (theta-1)*log(y) - theta*log(r) - y/r ;  
  model y ~ general(ll);  
  
run;
```

Zero-Inflated Gamma Model in PROC NLMIXED

The NLMIXED Procedure

NOTE: GCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood	12027
AIC (smaller is better)	12045
AICC (smaller is better)	12045
BIC (smaller is better)	12099

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
a0	0.8321	0.07224	2844	11.52	<.0001	0.05	0.6905	0.9738	-0.00003
a1	-0.2710	0.09241	2844	-2.93	0.0034	0.05	-0.4522	-0.08980	0.000031
a2	-0.2370	0.1247	2844	-1.90	0.0576	0.05	-0.4816	0.007653	0.000041
a3	-0.5772	0.1559	2844	-3.70	0.0002	0.05	-0.8829	-0.2715	0.000069
b0	3.2108	0.05292	2844	60.67	<.0001	0.05	3.1070	3.3146	-0.00005
b1	0.008780	0.06655	2844	0.13	0.8950	0.05	-0.1217	0.1393	0.00005
b2	0.1003	0.08907	2844	1.13	0.2602	0.05	-0.07435	0.2750	0.000067
b3	0.1164	0.1053	2844	1.11	0.2688	0.05	-0.08997	0.3228	0.000091
log_theta	0.2613	0.04044	2844	6.46	<.0001	0.05	0.1820	0.3405	7.42E-7

HLP ZILG Results Summary

- The ZILG results are the same as the results from the log-gamma and logistic models that were separately estimated. Why?
- Unlike a ZIP or ZINB model where zero is a possible value for the continuous part of the distribution, in the ZILG zero is not possible for the log-gamma part of the model. Whereas a response of zero in a ZIP or ZINB is a mixture of probabilities arising from being in the zero group vs. non-zero group **and** from being in the non-zero group, but having the possibility of having a zero response, in the ZILG the continuous and zero probability components of the distribution are completely separate.

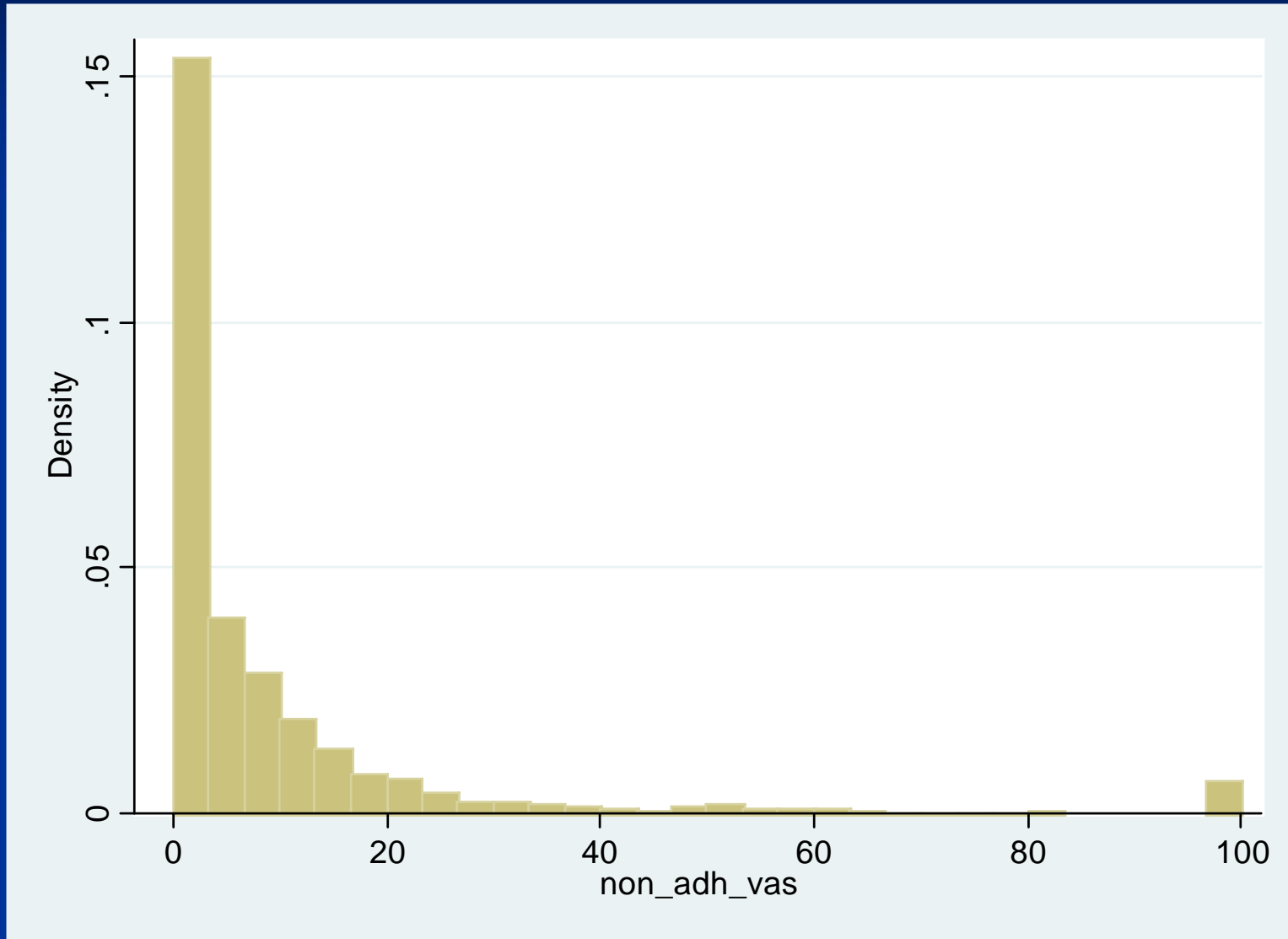
HLP ZIG Results Summary

- This means that for cross-sectional data there is little reason to model HIV medication non-adherence using a formal ZILG using NLMIXED – it is far more convenient to model the 0/1 and 1-or-more distributions separately using logistic and log-gamma regression, respectively.
- What about clustered or repeated measures data?
- The story may be somewhat different for clustered data.

Balance Project Example

- Balance is an intervention study designed to help persons with HIV cope with HIV symptoms and the side effects of ART (Mallory Johnson, PI)
- Two groups (intervention = 1 and control = 0) measured at five time points: baseline 1 (one week following screening), baseline 2 (month 3), post-intervention 1 (month 6), post-intervention 2 (month 9), post-intervention 3 (month 16)
- $N = 249$ who were randomized to intervention or control.

Distribution of Non-Adherence



Repeated Measures Non-Adherence

- As in the HLP example, there appears to be considerable zero-inflation of the non-adherence variable.
- As in the HLP example, break the problem down into two parts:
 - Random effects logistic regression of any non-adherence vs. perfect adherence
 - Random effects log-gamma regression for observations with 1% or more non-adherence.

Random Intercepts Logistic Model

```
proc glimmix data=temp
    method = quad (qpoints = 15)
    empirical = classical ;

class subject ;

model y_binary_vas(event='0') = randstat time group_by_time /
    solution
    dist = binary
    link = logit ;

random int / type = un subject = subject ;

nloptions tech = newwrap ;

run;
```

Random Intercepts Logistic Model

Fit Statistics

-2 Log Likelihood	794.96
AIC (smaller is better)	804.96
AICC (smaller is better)	805.01
BIC (smaller is better)	822.54
CAIC (smaller is better)	827.54
HQIC (smaller is better)	812.04

Covariance Parameter Estimates

Cov			Standard
Parm	Subject	Estimate	Error
UN(1,1)	SUBJECT	4.5446	1.1125

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-3.0905	0.3961	247	-7.80	<.0001
randstat	-0.3164	0.4788	946	-0.66	0.5089
time	-0.02959	0.03455	946	-0.86	0.3919
group_by_time	0.09210	0.04435	946	2.08	0.0381

Random Intercepts Logistic Model

- The significant group-by-time interaction indicates that the log odds of perfect adherence for the average intervention group participant is higher than that of the average control group participant over time.
- The estimate of the random intercept variance is four times larger than its standard error, which suggests the presence of considerable subject-specific variability in trajectories of adherence over time.

Random Intercepts Log-Gamma Model

```
proc glimmix data=temp
  method = quad (qpoinits = 15)
  empirical = classical ;
class subject ;
where y_binary_vas = 1 ;
model non_adh_vas = randstat time group_by_time /
                    solution dist = gamma link = log ;
random int / type = un subject = subject ;
nloptions tech = newrap ;
run;
```

Random Intercepts Log-Gamma Model

Fit Statistics

-2 Log Likelihood	6363.60
AIC (smaller is better)	6375.60
AICC (smaller is better)	6375.68
BIC (smaller is better)	6396.58
CAIC (smaller is better)	6402.58
HQIC (smaller is better)	6384.05

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	SUBJECT	1.5381	0.1490
Residual		0.7877	0.05454

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	1.6330	0.1222	242	13.37	<.0001
randstat	0.008406	0.1796	796	0.05	0.9627
time	0.01412	0.009269	796	1.52	0.1282
group_by_time	-0.00232	0.01428	796	-0.16	0.8711

Random Intercepts Log-Gamma Model

- No significant fixed effects are observed
- The random intercept parameter estimate is larger than its standard error, which suggests the possibility of non-trivial amounts of subject-specific variance in log-gamma trajectories of percent non-adherence
- Let's see what happens when we model both the logistic and log-gamma processes jointly via a random intercept zero-inflated log-gamma model (ZILG)
- As before, use PROC NLMIXED to fit the ZILG
 - Include random intercepts via the RANDOM statement
 - Request robust “sandwich” Huber-White standard errors via the EMPIRICAL option
 - Estimate the correlation of random intercepts via a custom ESTIMATE statement

Random Intercepts ZILG

```
proc nlmixed data = temp lognote = 3 qpoints = 15 tech = newwrap empirical ;  
  parms a0=0.1 a1=0.1 a2=0.1 a3=0.1 b0=0.4 b1=0.2 b2=0.4 b3=0.4 log_theta=0;  
  title "NLMIXED - Log-Gamma inflated model with repeated measures: Random  
intercepts" ;  
  y = non_adh_vas ;  
  
  linfp = a0 + a1*randstat + a2*time + a3*group_by_time + u1 ;  
  infprob = exp(linfp)/(1+exp(linfp)) ;  
  
  linp = b0 + b1*randstat + b2*time + b3*group_by_time + u2 ;  
  mu = exp(linp);  
  theta = exp(log_theta) ;  
  r = mu/theta ;  
  
  if y = 0 then ll = log(infprob) ;  
  else ll = log(1-infprob) - lgamma(theta) + (theta-1)*log(y) - theta*log(r) - y/r ;  
  model y ~ general(ll);  
  random u1 u2 ~ normal([0,0],[v11,c12,v22]) subject = subject ;  
  estimate "Corr(u1,u2)" c12 / (sqrt(v11*v22));  
  
run;
```

Random Intercepts ZILG

Fit Statistics

-2 Log Likelihood	7110.6
AIC (smaller is better)	7134.6
AICC (smaller is better)	7134.9
BIC (smaller is better)	7176.8

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
a0	-2.9827	0.3746	247	-7.96	<.0001	0.05	-3.7205	-2.2449	9.648E-6
a1	-0.4268	0.4656	247	-0.92	0.3602	0.05	-1.3437	0.4902	-6.24E-7
a2	-0.02921	0.03418	247	-0.85	0.3937	0.05	-0.09653	0.03811	-7.18E-6
a3	0.09071	0.04442	247	2.04	0.0422	0.05	0.003222	0.1782	-3.61E-6
b0	1.5549	0.1272	247	12.22	<.0001	0.05	1.3043	1.8055	-1.08E-6
b1	0.01907	0.1839	247	0.10	0.9175	0.05	-0.3431	0.3812	-2.95E-7
b2	0.01444	0.009317	247	1.55	0.1225	0.05	-0.00391	0.03279	-5.26E-6
b3	-0.00290	0.01430	247	-0.20	0.8394	0.05	-0.03106	0.02526	-2.6E-6
log_theta	0.2476	0.06884	247	3.60	0.0004	0.05	0.1120	0.3832	-1.06E-7
v11	4.2114	1.0482	247	4.02	<.0001	0.05	2.1468	6.2760	-0.00006
c12	-1.8062	0.3257	247	-5.55	<.0001	0.05	-2.4477	-1.1647	0.000011
v22	1.6688	0.1646	247	10.14	<.0001	0.05	1.3446	1.9930	-1.46E-6

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Corr(u1,u2)	-0.6813	0.08999	247	-7.57	<.0001	0.05	-0.8586	-0.5041

Random Intercepts ZILG

- Results are consistent with the previous random intercepts logistic and log-gamma models fitted with PROC GLIMMIX
 - The log odds of perfect adherence increase for the average intervention participant relative to the average control participant over time.
- There is a substantial negative correlation between the two random intercepts: the higher the odds of perfect adherence, the lower the value of the log-gamma intercept for non-adherence, which is conceptually sensible (if a participant is more likely to be perfectly adherent over time, the participant is less likely to have large values of non-adherence and vice versa).
- What about adding random slopes to these models?

Random Intercepts and Slopes Logistic

```
proc glimmix data=temp
    method = quad (qpoints = 15)
    empirical = classical ;

    class subject ;
    model y_binary_vas(event='0') = randstat time group_by_time
        / solution dist = binary
        link = logit ;
    random int time / type = un subject = subject ;
    nloptions tech = newrap ;

run;
```

Random Intercepts and Slopes Logistic

Fit Statistics

-2 Log Likelihood	794.41
AIC (smaller is better)	808.41
AICC (smaller is better)	808.51
BIC (smaller is better)	833.04
CAIC (smaller is better)	840.04
HQIC (smaller is better)	818.32

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	SUBJECT	4.5411	1.7037
UN(2,1)	SUBJECT	0.004274	0.1017
UN(2,2)	SUBJECT	0.006423	0.01015

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-3.0475	0.4830	247	-6.31	<.0001
randstat	-0.3303	0.4790	699	-0.69	0.4908
time	-0.05269	0.06354	247	-0.83	0.4077
group_by_time	0.09793	0.04733	699	2.07	0.0389

Random Intercepts and Slopes Logistic

- Results are similar to what we obtained in the previous analyses: a significant group-by-time interaction such that the average intervention participant has a higher log odds of being perfectly adherent when compared with the average control participant.
- There does not appear to be much variability in random slopes.

Random Intercepts and Slopes

Log-Gamma

```
proc glimmix data=temp
    method = quad (qpoints = 15)
    empirical = classical ;
class subject ;
where y_binary_vas = 1 ;
model non_adh_vas = randstat time group_by_time /
    solution dist = gamma link = log;
random int time / type = un subject = subject ;
nloptions tech = newwrap ;
run;
```

Random Intercepts and Slopes Log-Gamma

- Did not converge:

The GLIMMIX Procedure

NEWRAP needs more than 50 iterations or 500 function calls.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	SUBJECT	0.8115	.
UN(2,1)	SUBJECT	0.03265	.
UN(2,2)	SUBJECT	0.001314	.
Residual		1.6076	.

- Note how close the random slope estimate is to zero.
- This may be due to the relatively small number of cases with non-zero values for non-adherence (n = 932 out of 1198 observations had zero non-adherence).
- Modeling random slopes for those few cases may not be desirable.

Random Intercepts and Slopes ZILG

- So far our results suggest modeling the random intercepts for both the logistic and log-gamma model components, but not random slopes due to the non-significant slope estimate in the random intercepts-and-slopes logistic model and the non-convergent random intercepts-and-slopes log-gamma model. For illustrative purposes, however, let's see how to model random intercepts and slopes simultaneously using the ZILG.
 - Model random intercepts and slopes for the binary component of the model
 - Model random intercepts for the log-gamma part of the model
 - Model the covariance (correlation) between the two random intercepts and between the logistic random intercept and logistic random slope, but not between the log-gamma random intercept and the logistic slope
 - As before, use PROC NLMIXED
- Run time: 8 hours and 26 minutes
 - Tip: Use fewer quadrature points initially during testing

Random Intercepts and Slopes ZILG

```
proc nlmixed data = temp qpoints = 15 lognote = 3 tech = newrap empirical ;
  parms
    /* From random intercepts ZILG model */
    a0=-2.98 a1=-.43 a2=-.03 a3=.09
    b0=1.56 b1=.02 b2=.01 b3=-.003
    v11=4.21 c12=-1.81 v22=1.67 log_theta=.25
    /* From random intercepts and slopes logistic model */
    c13 = .04 v33 = .06
    ;
  title "NLMIXED - Log-Gamma inflated model with repeated measures: Random intercepts and
    slopes" ;
  y = non_adh_pct ;

  linfp = a0 + a1*randstat + a2*time + a3*group_by_time + u1 + u3*time ;
  infprob = exp(linfp)/(1+exp(linfp)) ;

  linp = b0 + b1*randstat + b2*time + b3*group_by_time + u2 ;
  mu = exp(linp);
  theta = exp(log_theta) ;
  r = mu/theta ;

  if y = 0 then ll = log(infprob) ;
  else ll = log(1-infprob) - lgamma(theta) + (theta-1)*log(y) - theta*log(r) - y/r ;

  model y ~ general(ll);

  random u1 u2 u3 ~ normal([0,0,0],[v11,c12,v22,c13,0,v33]) subject = subject ;
  estimate "Corr(u1,u2)" c12 / (sqrt(v11*v22));
  estimate "Corr(u1,u3)" c13 / (sqrt(v11*v33));
run;
```

Random Intercepts and Slopes ZILG

Fit Statistics

-2 Log Likelihood	3465.1
AIC (smaller is better)	3493.1
AICC (smaller is better)	3493.5
BIC (smaller is better)	3542.4

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
a0	2.7425	0.3867	246	7.09	<.0001	0.05	1.9808	3.5043	-0.00001
a1	-0.7147	0.4417	246	-1.62	0.1070	0.05	-1.5848	0.1554	5.379E-6
a2	-0.06837	0.03340	246	-2.05	0.0417	0.05	-0.1342	-0.00258	-0.00002
a3	0.06610	0.03805	246	1.74	0.0836	0.05	-0.00886	0.1411	-8.45E-6
b0	3.1206	0.1871	246	16.67	<.0001	0.05	2.7520	3.4892	2.879E-7
b1	0.08147	0.1847	246	0.44	0.6595	0.05	-0.2823	0.4453	-1.02E-6
b2	0.01480	0.01241	246	1.19	0.2342	0.05	-0.00964	0.03924	1.401E-6
b3	0.02177	0.01771	246	1.23	0.2200	0.05	-0.01310	0.05665	1.53E-6
v11	5.2984	1.7188	246	3.08	0.0023	0.05	1.9130	8.6837	-0.00012
c12	-0.2952	0.2700	246	-1.09	0.2753	0.05	-0.8271	0.2366	3.38E-6
v22	0.3643	0.07414	246	4.91	<.0001	0.05	0.2182	0.5103	-1.54E-6
log_theta	1.1323	0.1114	246	10.16	<.0001	0.05	0.9129	1.3517	2.423E-7
c13	-0.08235	0.09145	246	-0.90	0.3688	0.05	-0.2625	0.09778	8.896E-6
v33	0.008230	0.008323	246	0.99	0.3237	0.05	-0.00816	0.02462	-0.00003

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Corr(u1,u2)	-0.2125	0.1795	246	-1.18	0.2375	0.05	-0.5660	0.1410
Corr(u1,u3)	-0.3943	0.3773	246	-1.05	0.2970	0.05	-1.1375	0.3488

Random Intercepts and Slopes ZILG

- Variance of random slopes for the logistic part of the model and covariance of random slope with the logistic random intercept are not significantly different from zero.
- The random intercepts model may be better for this application, though the AIC and BIC favor the random intercepts-and-slopes model.
- The significant a_2 coefficient indicates that the log odds of perfect adherence for the average participant in the control group decreases over time.
- The logistic interaction effect, represented by coefficient a_3 , is marginally significant ($p = .084$), suggesting that intervention group participation could help the average participant attain perfect adherence more often over time.

Concluding Thoughts

- The log-gamma approach may be an appealing alternative to standard logistic regression methods for the analysis of HIV medication adherence data
- Log-gamma models for non-clustered cross-sectional data are readily available in Stata and SAS.
- Zero-inflated models may also be considered
 - Pros: May more accurately capture distributions with a preponderance of zeros
 - Cons: More complicated to set up, run, and interpret. Also, a lack of sufficient non-zero data may be concern.
 - For non-clustered data, splitting the data into zero vs. non-zero outcomes modeled with logistic regression and greater-than-zero data is a viable approach that is simpler to implement and yields identical parameter estimates and standard errors to the ZILG.

More Concluding Thoughts

- For repeated measures or clustered data, random effects ZILGs may yield similar, but not identical findings to separate random effects logistic and log-gamma models.
- It is unclear how much advantage is gained by using the random effects ZILG vs. separate logistic and log-gamma models.
- Other distributions can be explored (e.g., Log-normal), with or without zero-inflation considered using SAS, Stata, and other software programs.
 - There is a SAS macro, `mixcorr.sas`, written by Janet Tooze available for modeling zero-inflated normal and lognormal data with separate, but correlated random intercepts for the zero/not zero and the continuous not zero values.
 - Mplus has a nice set of features for two-part modeling of continuous data with a preponderance of zero values. In the Mplus approach, the non-zero components are treated as continuous and then transformed (the default transformation is the log, other transformations, including no transformation, are available).

Some Final Thoughts

Some comments from Chuck McCulloch:

- When one specifies a distribution (like Poisson) for a generalized linear model routine (glm or genmod) the program does not assume a Poisson distribution but instead garners two bits of information from that specification. First, it uses it to determine an optimal weighting scheme for calculating the coefficients. Second, it uses it to calculate model-based SEs, based on the distribution.
- Turning on the robust option (Stata; empirical in SAS) overrides the second aspect.
- In many applications, the weighting aspect is unimportant.
- So, if you are specifying the link and using robust SEs, a wide variety of different distributional choices will give the almost the same result.
- If you get far from the true distribution however, you can lose a bit of efficiency.

Acknowledgements

- Parya Saberi, PharmD
- Mallory Johnson, PhD
- Estie Hudes, PhD, MPH
- Eric Vittinghoff, PhD
- Chuck McCulloch, PhD
- Sam Dilworth, MS