How to Make a Caterpillar Plot with SAS Procs Glimmix and SGPlot (sometimes SurveySelect)

Brandy R. Sinco, BS, MA. MS

Outline

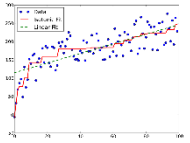
- 1) What is a Caterpillar Plot?
- 2) Proc Glimmix Model Syntax.
- 3) Process Fixed and Random Effects.
- 4) Combine Fixed and Random Effects for Each Facility.
- 5) Create the Caterpillar Plot.
- 6) When to Use Proc SurveySelect.
- 7) Closing Comments.

Caterpillar Plot

- Random effects model with facility_id as a random intercept. Capture natural variation by facility_id by including it as a random effect.

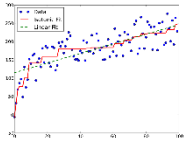
- Traditional mean is not accurate because the sum(X)/N is based on assumption that all X_i's are independent, when patients are clustered within facility.

- Also called empirical Bayesian estimate, reliability adjusted estimate, best linear unbiased predictor (BLUP).
- Equivalent to empirical Bayesian estimate if prior distribution of random effects are assumed to be normal (Robinson, 1991).



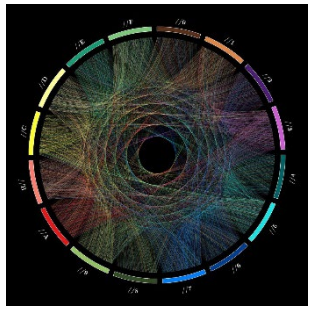
Linear Mixed Model

- Y = outcome, for which the error distribution in a regression model, ε, is normal.
- The equation for the linear mixed model is: $Y = X\beta + Zb + \varepsilon$.
- Instead of $\varepsilon \sim N(0, \sigma^2)$ in linear regression, $\varepsilon \sim N(0,\Sigma)$ because the residuals can be correlated in a LMM.
- ε is always a random effect; there is an estimate for each subject.
- $b \sim N(0, G)$, where G = covariance matrix of the random effects, other than ε.
- **Finally, G and Σ are assumed to be independent.**
- **Var(Y) = Var(**Zb**) + Var(**ε**).**



Generalized Linear Mixed Model (GLMM)

- Used to model a binary or count outcome, where the error distribution is not necessarily normal. The outcome, Y, must be from an exponential family.
- Model the mean of Y as a non-linear function of ($X\beta + Zb$), called the link function.
- For binary Y with probability. π, use logit(π) as the link function, where $\text{logit}(\pi) = \ln(\pi/(1 - \pi))$.
- $\text{logit}(\pi) = \ln(\pi/(1 - \pi) = X\beta + Zb$.
- To estimate, π, proportion of event outcome, $\pi = \exp(X\beta + Zb)/(1+\exp(X\beta + Zb))$.



Reliability Adjusted Procedure Rate at Facilities

- π_{ij} = probability of procedure for i^{th} woman at j^{th} facility.
- β_0 = population average intercept for all women, all facilities.
- b_{j0} = intercept for the j^{th} facility (also known as random intercept).
- $\text{Logit}(\pi_{ij}) = \beta_0 + b_{j0.}$
- Procedure rate at jth facility, $\pi_{ij} = \exp(\beta_0 + b_{j0}) / (1 + \exp(\beta_0 + b_{j0}))$.
- Percentages at jth facility = $100\pi_{ij} = 100\exp(\beta_0 + b_{j0}) / (1 + \exp(\beta_0 + b_{j0}))$.
- Reference: Dimick, et al, 2012. Reliability Adjustment for Reporting Hospital Outcomes With Surgery.

Use SAS Proc Glimmix to Obtain π_{ij}

```
ods html path="c:\temp"; ods graphics on;

/* method=quad uses Gauss-Hermite quadrature instead of default rspl */
proc glimmix data=margins_nomiss method=quad plots=all NOCLPrint;

  class PUF_FACILITY_ID;

/* ReOperationN: 0 = no, -1 = yes, reference largest value 0 */
  model ReOperationN = / dist=binary  link=logit solution;
  random int/subject=PUF_FACILITY_ID/ Solution; covtest 0;

Ods Output ParameterEstimates=ReOp_FixedEffects
SolutionR=ReOp_RandEffects; /* output fixed & random effects */

run; ods graphics off; ods html close;
```

Glimmix Output

Fixed Effects:

Effect	Estimate	Std Err	DF	t Value	Pr > |t|
Intercept	-1.56	0.02	1213	-89.06	<.0001

Facility Estimates $\{b_{0j}\}$:

Subject	Estimate	Std Err Pred	DF	t Value	Pr > |t|
PUF_FACILITY_ID AAHNYLQGYK	-0.38	0.25	1.39E+05	-1.49	0.1371
PUF_FACILITY_ID AARCBCDTMV	-0.02	0.22	1.39E+05	-0.1	0.922
PUF_FACILITY_ID AASXTTOARC	-0.49	0.36	1.39E+05	-1.37	0.171
PUF_FACILITY_ID AAXODGTRTR	1.03	0.18	1.39E+05	5.82	<.0001

Fixed Intercept, β_0

```
/* Process Fixed Effect Estimates By Facility */
Data Margins_Estimates_Fixed;
set ReOp_FixedEffects(keep=Effect Estimate StdErr);
rename Estimate=Estimate_Fixed;
rename StdErr=StdErr_Fixed;
Run;

/* Store fixed intercept and standard error in macro variables */
Data _NULL_;
set  Margins_Estimates_Fixed;
call symput('Estimate_Fixed_Margins',Estimate_Fixed);
call symput('StdErr_Fixed_Margins', StdErr_Fixed);
run;
```

Random Intercepts, b_{0j}

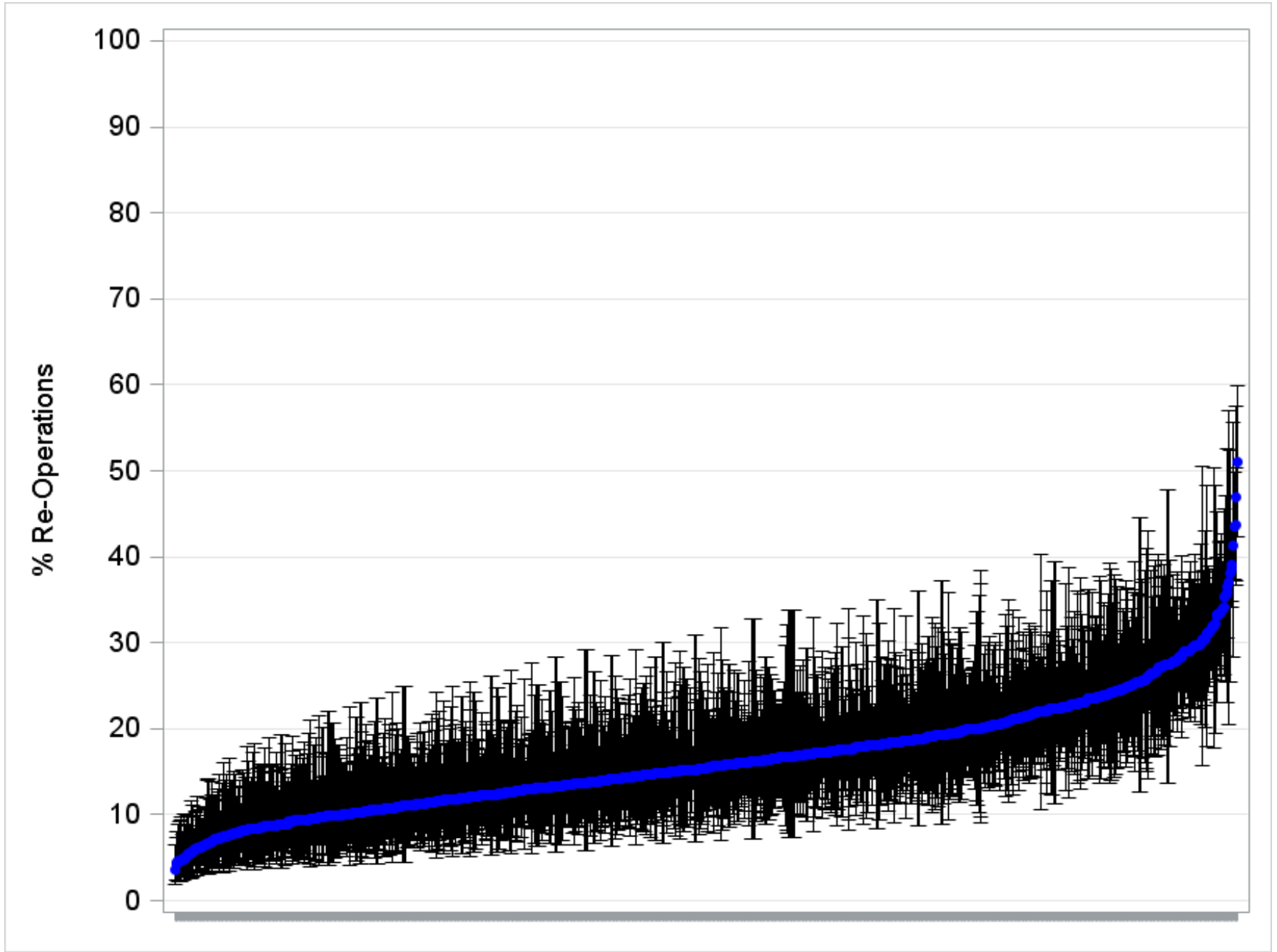
```
/* Process Random Effect Estimates by Facility */
Data Margins_Estimates_Random;
set ReOp_RandEffects(keep=Subject Estimate StdErrPred);
Format PUF_FACILITY_ID $10.;
rename Estimate=Estimate_Random;
rename StdErrPred=StdErr_Random;

/* Retrieve Facility ID From String Functions */
SpaceLocn=Find(Subject, " ", 1);
LenSubject=LengthC(Subject);
PUF_FACILITY_ID=Substr(Subject,SpaceLocn+1, LenSubject-SpaceLocn);
Run;
```

Compute Reliability-Adjusted Procedure Rates

```
Data Margins_Estimates_Combined;
Set Margins_Estimates_Random(Keep=PUF_FACILITY_ID Estimate_Random
StdErr_Random);

Estimate_Fixed= &Estimate_Fixed_Margins;
StdErr_Fixed= &StdErr_Fixed_Margins;

Estimate_Combined = Estimate_Fixed + Estimate_Random;
StdErr_Combined = sqrt((StdErr_Random**2)+(StdErr_Fixed**2));
LCL=Estimate_Combined-(1.96*StdErr_Combined);
UCL=Estimate_Combined+(1.96*StdErr_Combined);

Percent_Margins=100*exp(Estimate_Combined)/(1+exp(Estimate_Combined));
Percent_Margins_LCL=100*exp(LCL)/(1+exp(LCL));
Percent_Margins_UCL=100*exp(UCL)/(1+exp(UCL));
Run;
```

Caterpillar Plot

```
/* Randomly select 1000 of 1151 points */
proc surveyselect data=Brandy.Margins_Estimates_Combined method=srs
out=Margins_Estimates_Combined seed=5062020 n=1000;  run;

proc sgplot data=Margins_Estimates_Combined NoAutoLegend;

 scatter x=PUF_FACILITY_ID y=Percent_Margins /
ERRORBARATTRS=(color=black) yerrorlower=Percent_Margins_LCL
yerrorupper=Percent_Margins_UCL
        markerattrs=(symbol=circlefilled color=blue size=12);
XAXIS label="Facility" DISPLAY=(NOVALUES) LABELATTRS=(Size=12) ;
yaxis grid label='% Re-Operations' values=(0 to 100 by 10)
LABELATTRS=(Size=12) VALUEATTRS=(Size=12);
run;
```


100
90
80
70
60
50
40
30
20
10
0
% Re-Operations

How to Create .svg (Scalable Vector Graphic for JAMA)

```
ods html5 options (svg_mode="inline");
ods graphics /imagefmt=svg;
proc sgplot data=Adjuvant_Estimates_Combined NoAutoLegend;
scatter x=site_code y=Percent_Adjuvant /  ERRORBARATTRS=(color=black)
yerrorlower=Percent_Adjuvant_LCL yerrorupper=Percent_Adjuvant_UCL
  markerattrs=(symbol=circlefilled color=red size=12);
XAXIS label="Facility" DISPLAY=(NOVALUES) LABELATTRS=(Size=12) ;
yaxis grid label='% Adjuvant' values=(0 to 100 by 10)
LABELATTRS=(Size=12) VALUEATTRS=(Size=12);
run;
ods graphic off;  ods html5 close;
```

References

- Robinson GK. “That BLUP is a Good Thing: The Estimation of Random Effects,” Statistical science, vol. 6, no. 1, pp. 15–32, 1991.

- Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall, BL, “Reliability adjustment for reporting hospital outcomes with surgery,” Annals of surgery, vol. 255, no. 4, pp. 703–707, 2012.

- Wang T, Bredbeck BC, Sinco B, Shubeck S, Baskin B, Skolarus T., Dossett LA. Variations in Persistent Use of Low-Value Breast Cancer Surgery. JAMA Surg. Published online February 03, 2021. doi:10.1001/jamasurg.2020.6942.

Closing Comments

- SAS Proc Glimmix is used to create caterpillar plot dataset with facility_id as random effect. The output from Proc Glimmix can be routed to graphics procedures, such as Proc SGPlot for caterpillar plots, histograms, box plots, and odds ratio plots.

- Proc SGPlot can easily be modified to create scalable vector graphics (.svg) files for medical journals.

- Proc SGPlot can create a caterpillar figure with the scatter command. The 95% confidence intervals can be displayed with error bars. In Proc SGPlot, need to suppress display of the facility id's with DISPLAY=(NOVALUES)

- If the data contains more than 1,000 facilities, randomly select 1,000 with Proc SurveySelect before plotting the data with Proc SGPlot.

Acknowledgements

- Drs. Ton Wang, Brooke Bredbeck, Sarah Shubeck, Alison Baskin, Ted Skolarus, Lesly A. Dossett.

- Dr. Wang is supported by a grant from the National Cancer Institute (NCI; T32 CA009672), Dr. Bredbeck is supported by the Ruth L. Kirschstein Research Service Award from the National Cancer Institute (NCI; T32 CA009672), Dr. Skolarus is supported by grants from the National Cancer Institute (R01 CA242559 and R37 CA222885), and Dr. Dossett is supported by a grant from the Agency for Healthcare Research and Quality (AHRQ; 5 K08 HS026030-02).

Contact Information

Brandy R. Sinco, BS, MA, MS, Statistician Senior
Michigan Medicine
Center for Healthcare Outcomes and Policy
2800 Plymouth Rd.
Ann Arbor, MI 48109-2800

E-Mail: brsinco@med.umich.edu