Generalized Linear Mixed Model Approach to Time-to-Event Data with Censored Observations

K. Yeater[1], G. Yocum[1], J. Rinehart[1], A. Rajamohan[1], J. Bowsher[2], and K. Greenlee[2]

[1]USDA-ARS-PA-ETSARC-Insect Genetics and Biochemistry Research, Fargo, ND

[2]Dept. of Biological Sciences, North Dakota State Univ., Fargo, ND

NDSU USDA ARS

Time-to-Event : Survival Analysis Methods

- *Survival Analysis* is a class of methods for which the outcome variable of interest is time until an event occurs.
- Time is measured from beginning (time=0) until the event occurs or the observation time ends.

Goals of Survival Analysis

- Estimate and interpret survival
- Compare survival among different groups
- Assess relationship of time-independent and time-dependent explanatory variables
- Predict time until the event

Conventional Stat Methods

- Logistic Regression
  - Ignores timing of events
  - Cannot handle time-dependent variables
- Linear Regression
  - Cannot handle censored observations
  - Cannot handle time-dependent variables
  - Time to event can have non-normal distribution

Uninformative vs. Informative Censoring

- Uninformative if it occurs when the reasons for removal are unrelated to the event and does not bias the parameter estimates and statistical inference
- Informative if it occurs when the reasons for removal are related to the event

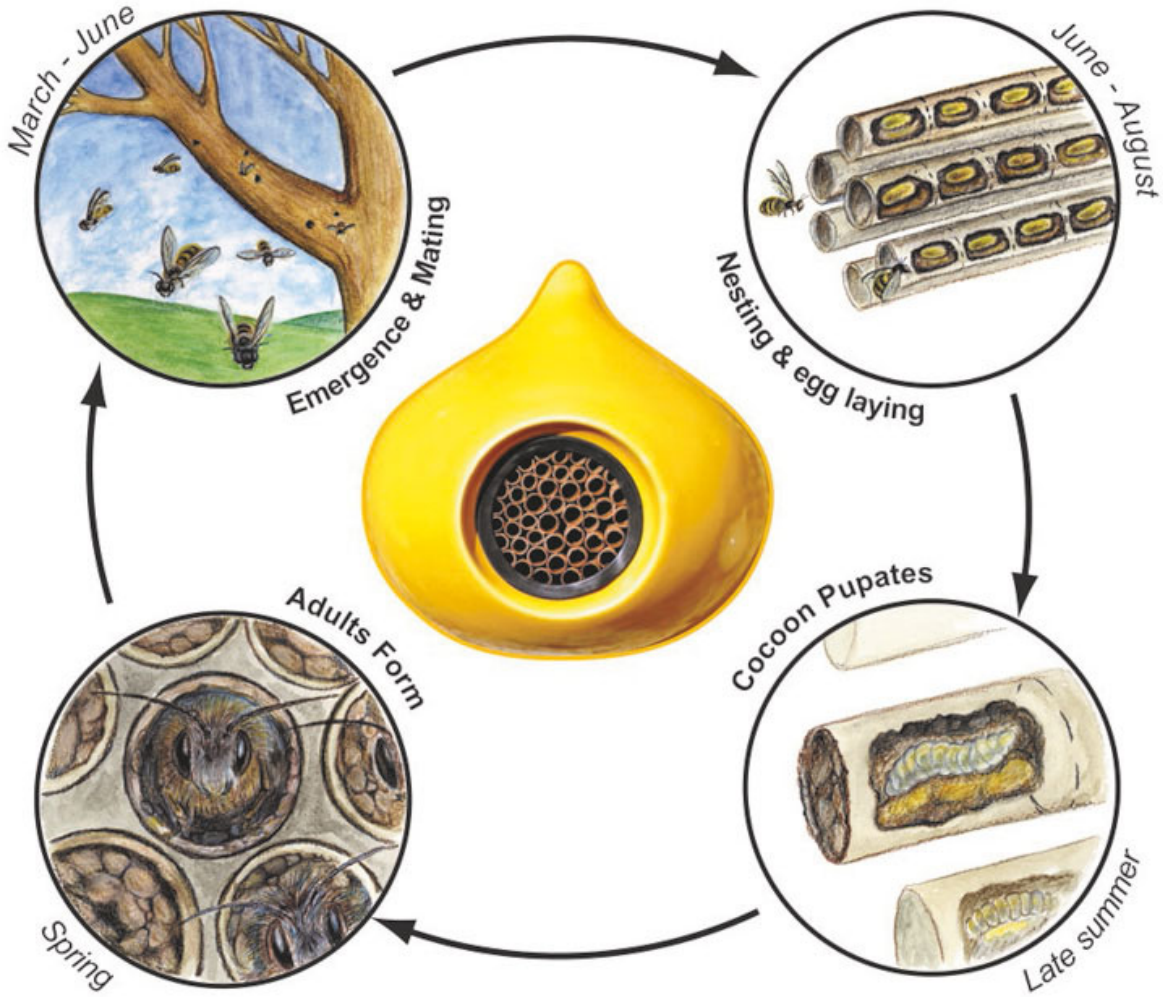
Survival / GLMM Connection

- Survival analysis
  - Estimate the survivor / hazard function
  - Other inferences of interest tied to these functions
- GLMM
  - Treatment and experiment designs – adding another layer
  - Survival fits nicely with a GLMM using the exponential distribution – which is based on the Poisson process

Research Problem

- Alfalfa leafcutter bees (*Megachile rontundata)*
- Pollinator management in Northern Plains and Canada
- Alfalfa, Canola, Camelina, Legumes

Life-cycle for ‘solitary’ bees



Beepalace.com/about_bee

Life Cycle and Management

Study Design

Developmental Stage

Identify, Categorize Pre-Pupae

Emergence Ready

Eye Pigment

Randomized Treatments Thermoprofile (2 Controls and 8 Temperature Regimes) X Storage Time (weeks 1 to 8)

3 x 24 cells (72 obs) per Therm x Week

Emergence Stage

Incubating Environment 29° C

Observe for ~40 days

Exponential Survival GLMM (for Censored Data)

- Treatment Structure – Treatment X Week Factors (Fixed Effects)
- Design Structure – w/in each Trt x Wk combination, plates observed (Random Effects)
- 1st Response Variable – Days to Emergence (our “survival” time)
- 2nd Response Variable – Censored / *not* Censored (binary)

Censored Observations

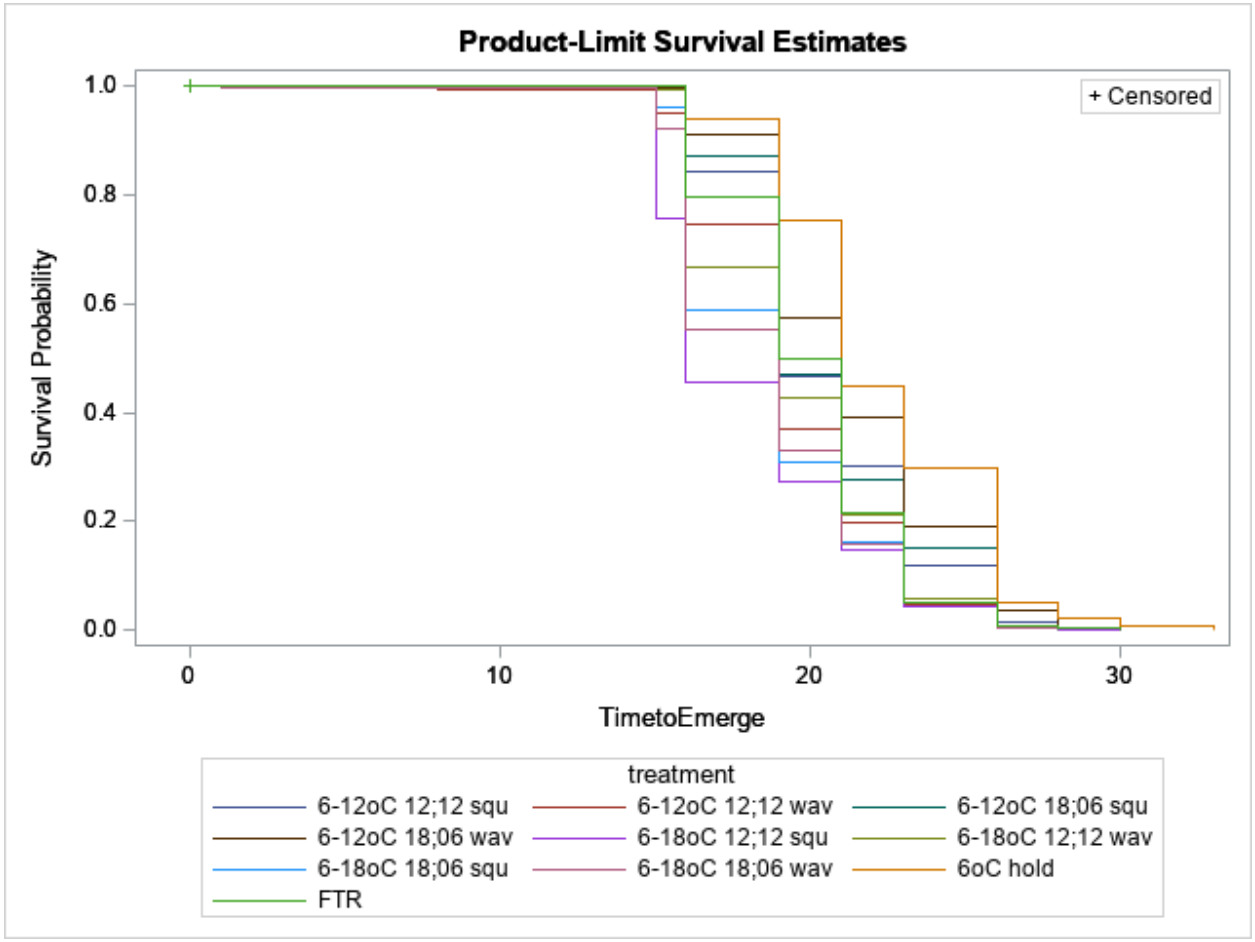
- If an obs is *censored*, it did not emerge prior to end of study (informative) – indicative of death or wasp emergence
- If an obs is *not censored*, the observed bee emerged while at 29°C, prior to conclusion of study

Data Exploration - % Censored

Development stage = Emergence Ready

Summary of the Number of Censored and Uncensored Values					
Stratum	treatment	Total	Failed	Censored	Percent Censored
1	6-12oC 12;12 squ	576	245	331	57.47
2	6-12oC 12;12 wav	576	244	332	57.64
3	6-12oC 18;06 squ	576	232	344	59.72
4	6-12oC 18;06 wav	576	209	367	63.72
5	6-18oC 12;12 squ	576	458	118	20.49
6	6-18oC 12;12 wav	576	350	226	39.24
7	6-18oC 18;06 squ	576	390	186	32.29
8	6-18oC 18;06 wav	576	398	178	30.90
9	6oC hold	576	134	442	76.74
10	FTR	576	273	303	52.60
Total		5760	2933	2827	49.08

Data Exploration – Kaplan-Meier



Exponential Model

- In the exponential model, the resulting log-likelihood is the Poisson.
- The censoring random variable C has a Poisson distribution with rate parameter λ.
- At given time t, $E(C \mid t) = \mu_C = \lambda t$

GLMM model

- Use *C* as the response variable
- Poisson as the conditional distribution of *C* given the random model effects
- $\text{Log}(\mu_C) = \log(\lambda) + \log(t)$ is the link
- Because the form of the log-likelihood works w/in the GLMM estimating equations, the model uses log(*t*) as an *offset*

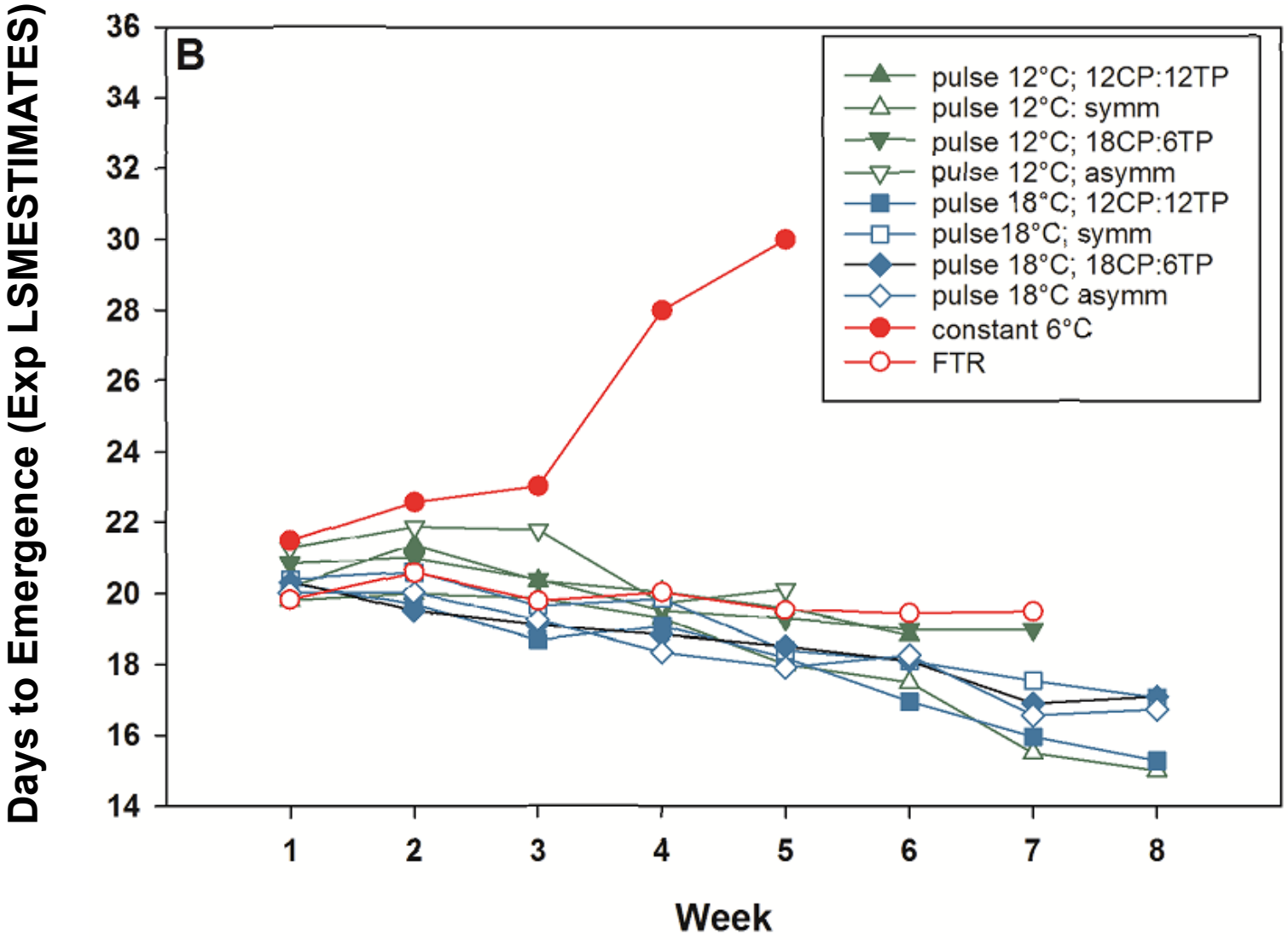
Stroup, Walter W. (2013) Generalized Linear Mixed Models: Modern Concepts, Methods and Applications. CRC Press, Boca Raton, FL

SAS GLIMMIX Proc Statements

- Model c = trt*wk / d=poi offset=logt;
- random int plate / subject=plate(trt*wk);
- lsmeans trt*wk / ilink;
- lsmestimate trt*wk 'mean emerge_time for 6-12oC 12;12 squ 1' -1 / exp;

LSMEANS/LSMESTIMATES

- LSMEANS provides estimates of log(λ) – rate parameter
  - ILINK gives actual values of the estimated hazard function for each TRT x WK
- LSMESTIMATES / EXP allow us to obtain exponentiated estimates of mean survival time – interpreted as mean “emergence”



Emergence Ready Development Stage

What's the Difference?

- GLMM vs Proportional Hazards vs Kaplan Meier….
- Now including random model effects
  - Exist as consequence of study design
  - Important to account for them, or we raise same conditional/marginal model issues, standard error issues, test statistic issues that we see for other mixed models with non-Gaussian data.