A statistical framework for analyzing dose-response repellent data and related issues

Matt Kramer

matt.kramer@ars.usda.gov

Biometrical Consulting Service, ARS/BARC/USDA

Outline

- General framework for dose-response model
- Model components
- EDA$_{50}$, EDA$_{95}$, sample size
- Examples
- Field tests
- Comparison of repellents
- CPT (TFB) measure and alternatives
- Power

General framework

- Model based on binomial response variable (known n)
- fixed effects: compound, concentration, species, ???
- random effects (good representation is important)
  - volunteer
  - block—these are variable(s) that group mosquitoes, e.g. tested in same day or in same cubicle, may be nested in volunteer or crossed
  - error

Response variable

- an individual mosquito
- a group of mosquitoes (e.g. 5/20 bit)
- my preference for a transformation of p is the logit (p) = log $\frac{p}{1-p}$, where p is the "true" proportion of mosquitoes biting
- logit (p) can be used if data is collected on individual or grouped mosquitoes
- *generalized* linear model because p is not modeled directly—model a function of p, and we expect p to follow a binomial distribution (parameters of binomial distribution affected by concentration, formulation, etc.)

Concentration (a fixed, independent variable)

Concentration

- often is linear with logit (p)
- if not, try a transformation, e.g. log, squareroot
- polynomial less desirable as interpretation is not clear

Example model

logit $(p) = \mu\ (+T_i) + \beta_1\sqrt{C_j} + \beta_i T_i\sqrt{C_j} + V_k + B_l$

- fixed effects: T_i, C_j
- random effects: V_k, B_l
- T_i by itself should not be needed (i.e. a different intercept for each treatment)—interest is in whether treatment slopes differ $(T_i\sqrt{C_j})$
- Since there are random effects, this is now a *generalized* linear *mixed* model

Estimating EDA$_{50}$, EDA$_{95}$

- This model can be used to estimate EDA$_{50}$, EDA$_{95}$
- You are asking the question: at what concentration is logit $(p) = 0$ (EDA$_{50}$)?
- Confidence interval estimation involves *fiducial limits*
- some programs produce this (and 95% confidence limits), but not for genralized linear mixed models
- for generalized linear mixed models, easy to get a point estimate (e.g., the EDA$_{50}$), but not its variance.

Notes on this type of model

- Each volunteer gets at least 2 treatments, more if possible (look at incomplete block designs where each volunteer is a block)
- Assumption that individual mosquitoes are independent. If there are correlated responses not accounted for in the model, either missing something in the model or there is a problem with the methods/technique
- If the B_l term(s) is(are) ignored, the response will seem overdispersed compared to a binomial distribution (may called quasi-binomial distribution in a statistical program)
- What concentrations to use? Depends on what you want to estimate. EDA$_{50}$? EDA$_{95}$? Put concentrations around these points (need preliminary trials for this). Whole curve? Need lots of concentrations. Why would you be interested in this? Mimic a time (wearing off) effect?

Low concentrations have no effect

- curve really starts at a low, but nonzero concentration of repellent
- subtract out the beginning flat part of the curve
- example: if curve starts at a concentration of 0.05, use $\beta_1\sqrt{C_j - 0.05}$ when estimating the model and ignore data at $C_j < 0.05$

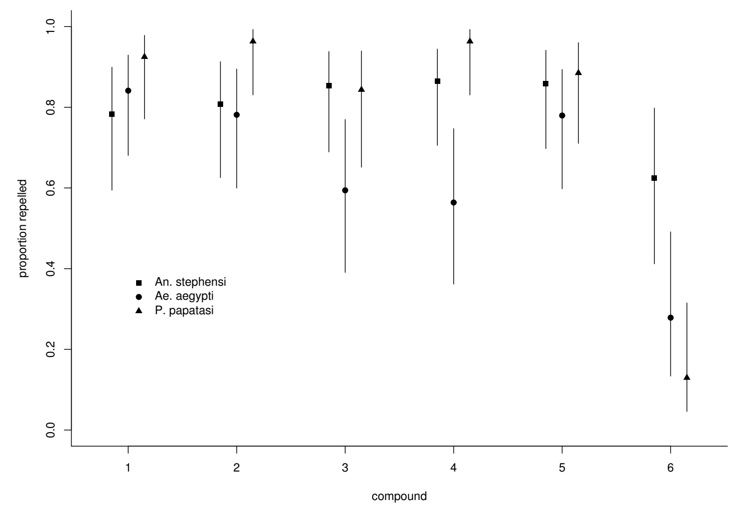
Sample size in dose response

- binomial data not very informative compared to quantitative data
- need lots of mosquitoes (but mosquitoes are cheap)
- put most mosquitoes at dose(s) you are most interested in (probably high concentrations)
- *not* necessary to distribute mosquitoes evenly across concentrations

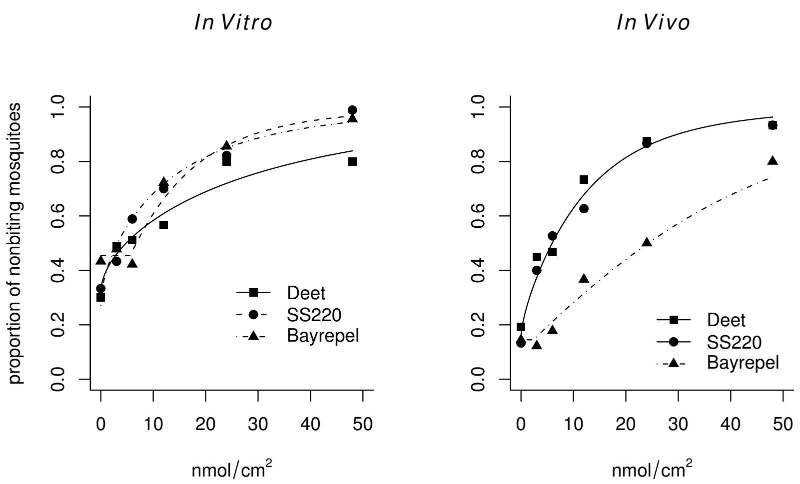
Sample size in dose response

- rule of thumb—need about 50 mosquitoes to get within a $\sim$10% accuracy
- standard deviation $= \sqrt{\frac{p(1-p)}{n}}$
- $\Rightarrow n = \frac{p(1-p)}{SD^2}$
- n will be greatest for $p = 1 - p = 0.5$
- Example: want 10% accuracy at $p = 0.9$
- 2 $\times$ SD = 0.10 (i.e. $\sim$ coverage of 95%) $\Rightarrow$ use SD = 0.05
- $n = \frac{p(1-p)}{SD^2} = \frac{0.9\times 0.1}{0.05^2} = 36$ mosquitoes per treatment-dose combination

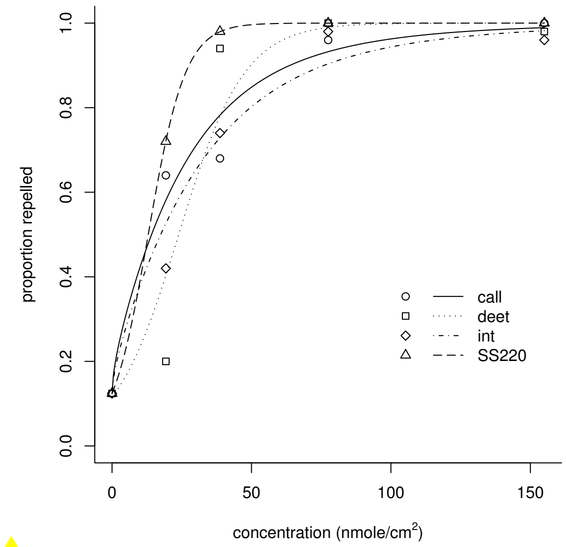
Example



Example



Example



Field Studies (n is unknown)

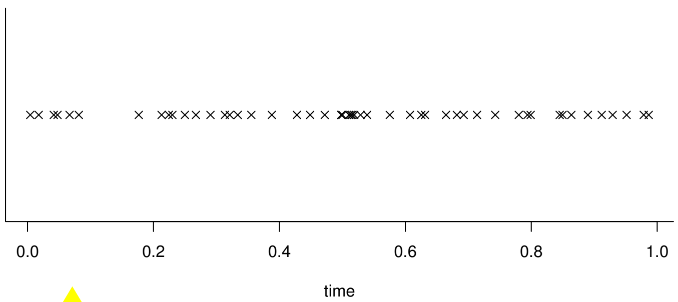
- more difficult because number of mosquitoes is unknown
- need to have a measure of biting pressure
- need to know that n is large
- need controls (volunteers) that are susceptible to mosquito bites
- design experiment so each volunteer serves as his/her own control
- need a lot of volunteers to control for (estimate) person-to-person variability—also to make sure repellent is effective across the population
- best to use same volunteer trying several different formulations, may have to block by days (so look at blocking designs)

Field Studies (n is unknown)

- use concentrations where mosquitoes bite, otherwise formulations will appear to be the same
- hard to estimate at what concentration two different formulations are the same, better to ask if formulations differ at same concentration
- cannot use binomial response model because n is unknown, consider Poisson model (count number of biting mosquitoes). This is still a generalized linear mixed model but assumes a different distribution for the response variable
- don't use χ^2 tests or other contingency table methods because other effects in experiment are ignored (assumption that mosquitoes are independent is not met, p values will be wrong)

CPT/TFB

- this statistic has poor properties (results depends on one mosquito!)
- example—assume after the repellent wears off (label this time 0), distribution of bites follows a uniform distribution
- i.e. time between bites is random until last mosquito that is going to bite does bite



CPT/TFB—assume bites follow a uniform distribution

- Expected value to TFB = $\frac{1}{n+1}$ (of course, we didn't know n to begin with)
- Variance of TFB = $\frac{n}{(n+1)^2(n+2)}$, so the SE $= \sqrt{\frac{n}{(n+1)^2(n+2)}}$
- example $n = 10$: expected value $= \frac{1}{11} = 0.91$
- if the test lasts 30 min after true CPT, then $30 \times 0.91 = 2.73$ min
- SE $= 2.49$, so a 95% CI $\sim (-2.25, 7.71)$
- example $n = 100$: expected value $= \frac{1}{101} = 0.0099$
- if the test lasts 30 min after true CPT, then $30 \times 0.0099 = .297$ min
- SE $= 0.294$, so a 95% CI $\sim (-0.291, 0.885)$

CPT/TFB

- Conclusion: number of mosquitoes that will bite (n), which is unknown, will determine results of the test
- Need to use a very large number of mosquitoes to be sure that the first bite occurs near the end of the protection time for that individual
- Might use the same (large) number of mosquitoes several times (perhaps even mark the first few that bite) to learn something about the process (how repeatable are the TFBs, is it always the same mosquitoes that bite first, how do things change if a new volunteer or formulation is used)

Alternatives to CPT/TFB

- use the mean of the time to the first five bites. SE will decrease as $\sim \frac{1}{\sqrt{n}}$, so $\frac{1}{\sqrt{5}} \approx \frac{1}{2}$
  - similar to the power increase from using 5 $\times$ as many mosquitoes
  - problem—you are no longer estimating protection time, however, it would be a much better way of detecting formulation differences near CPT
- repeat TFB may times per volunteer
  - problem—requires lots of trials per volunteer

Alternatives to CPT/TFB

- find true distribution of biting times (say, using membranes, not people!), and use that to project (forward) the end of CPT based on the first few bites of volunteers
    - problem—requires preliminary testing, the belief that results from membranes transfer over to humans, and some math to do the estimation, readers have to accept methodology.
    - However, this would probably provide the best estimate of the true CPT.

Power

- Easy to figure out for simple experiments (based on binomial distribution), available in many software packages
- Hard to calculate for generalized linear mixed models. I use Monte-Carlo (simulation based) methods
- Software: SAS or R to estimate generalized linear mixed models. R also has a number of packages specifically for dose-response, but these don't allow for random effects.

THE END