

Using Count-Based Statistical Models for Key Performance Indicators, Risk Metrics, *etc.*

Aaron J. Mackey, PhD
Director Data Sciences
Covance Informatics

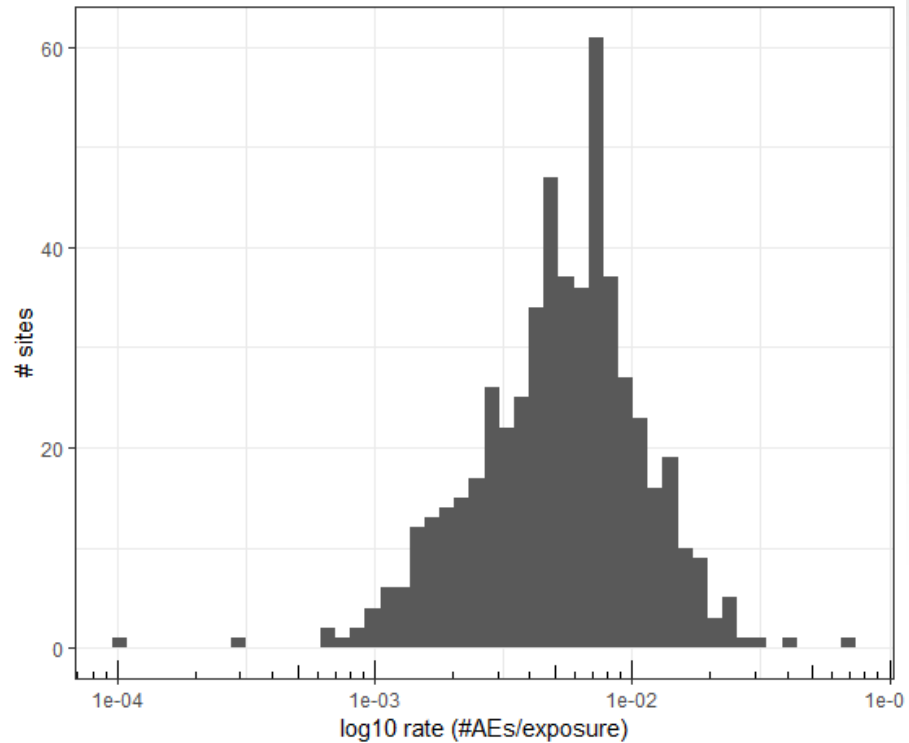
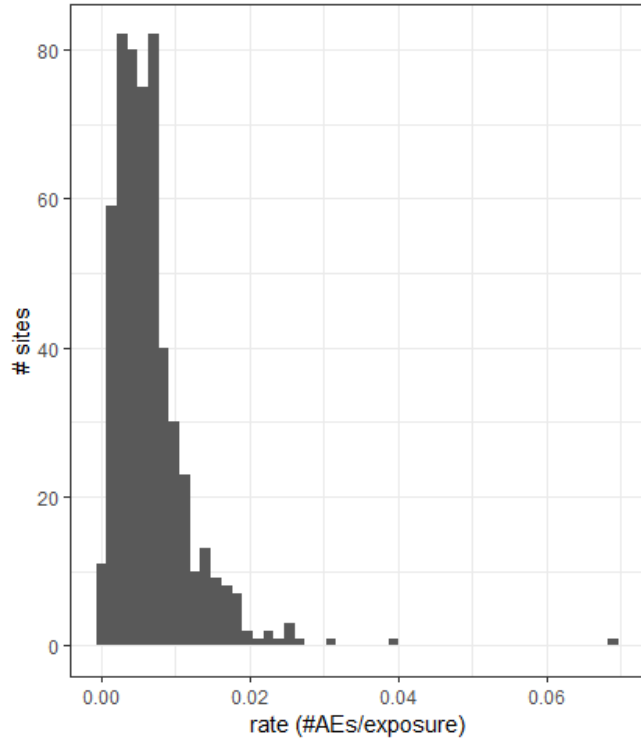
PhUSE US Connect 2019



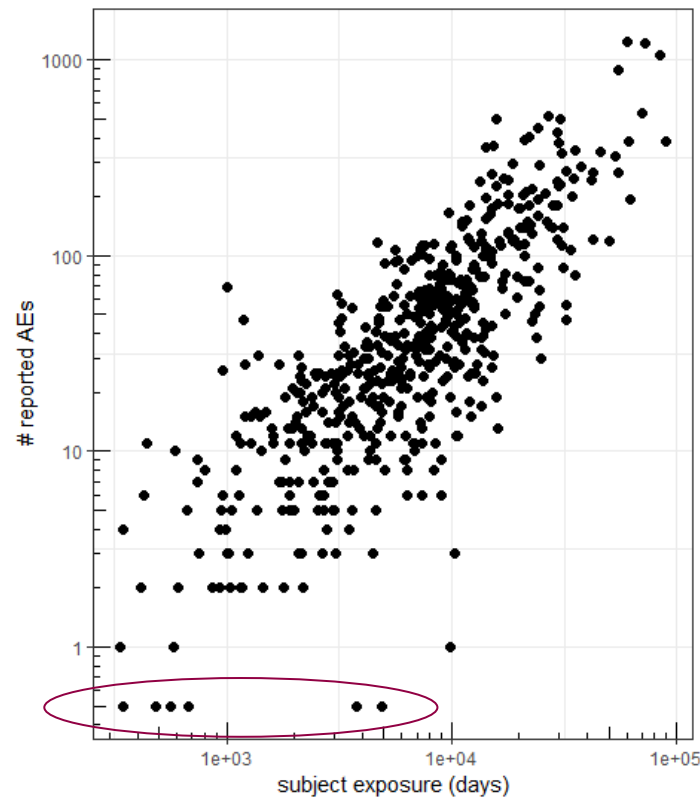
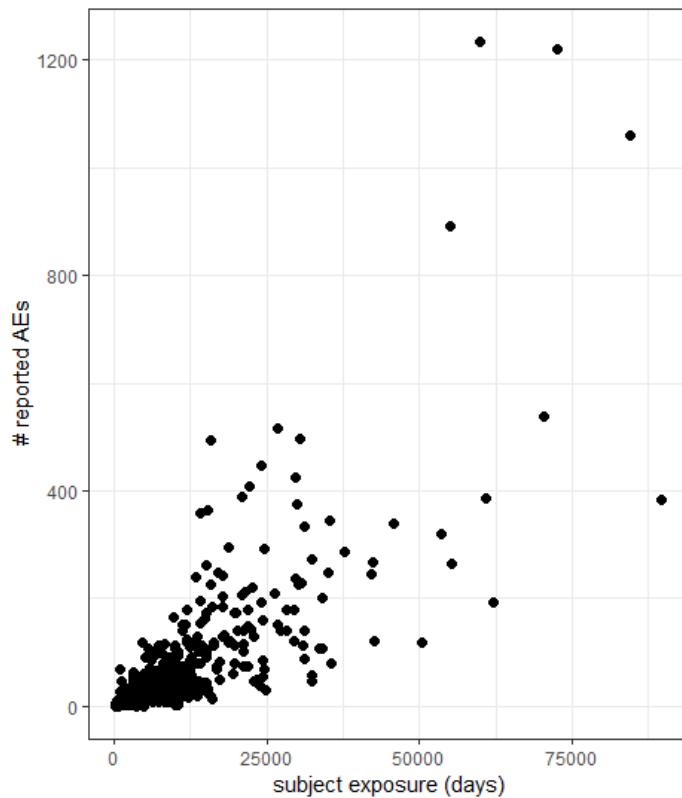
Why counts?

- ▶ Clinical trials are an amalgam of events, occurring in time
 - Country- and site-level activation, monitoring, close
 - Recruitment, Screening, Enrollment
 - Subject Visits
 - Adverse Event (AE) reports
 - Protocol Deviations, *etc.*
- ▶ Many KRI/KPI/QTLs are calculated as rates: counts / timespan
 - == mean expected counts of a *homogenous* Poisson process
 - Rates cannot be negative; thus *never* Normal-distributed
 - So why are so many statistics based on Z-scores of rates?
 - What better statistical distributions/models are available?

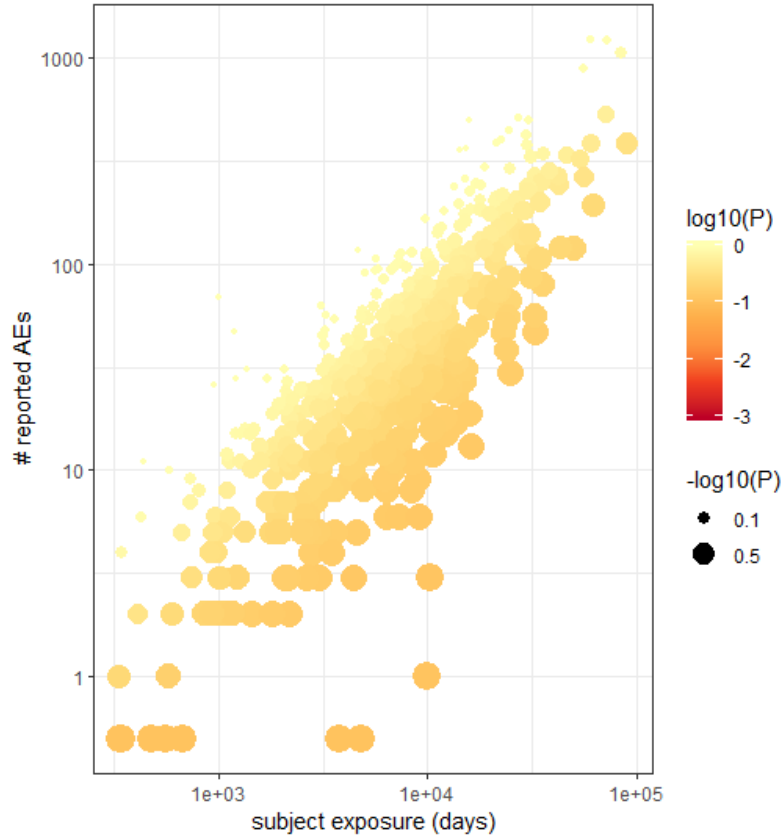
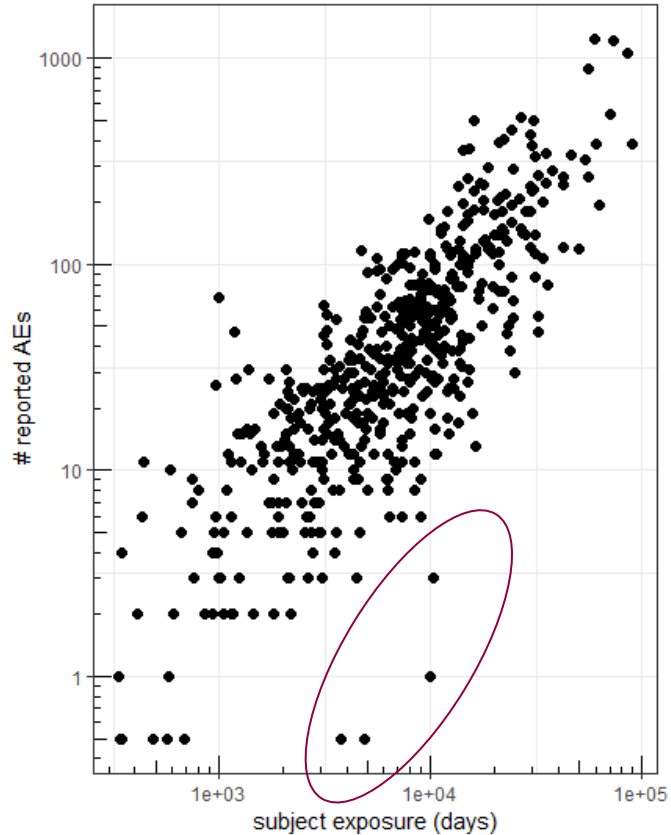
KRI: Adverse Event rates, under-reporting sites?



Linear vs. Logarithmic scales: watch the zeroes!



the Normal/Z-score treatment: no sites significant?

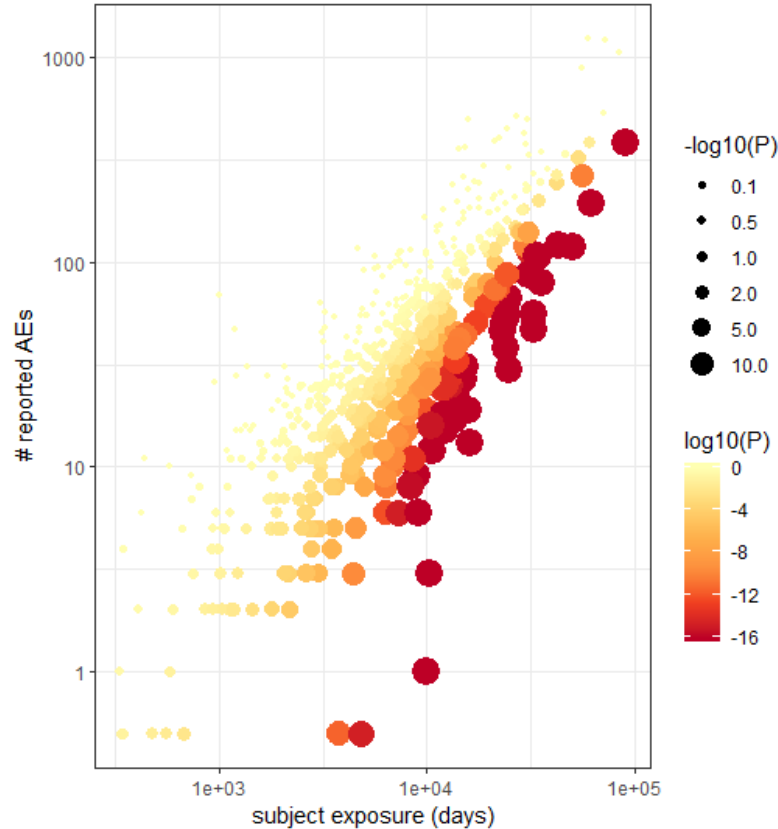
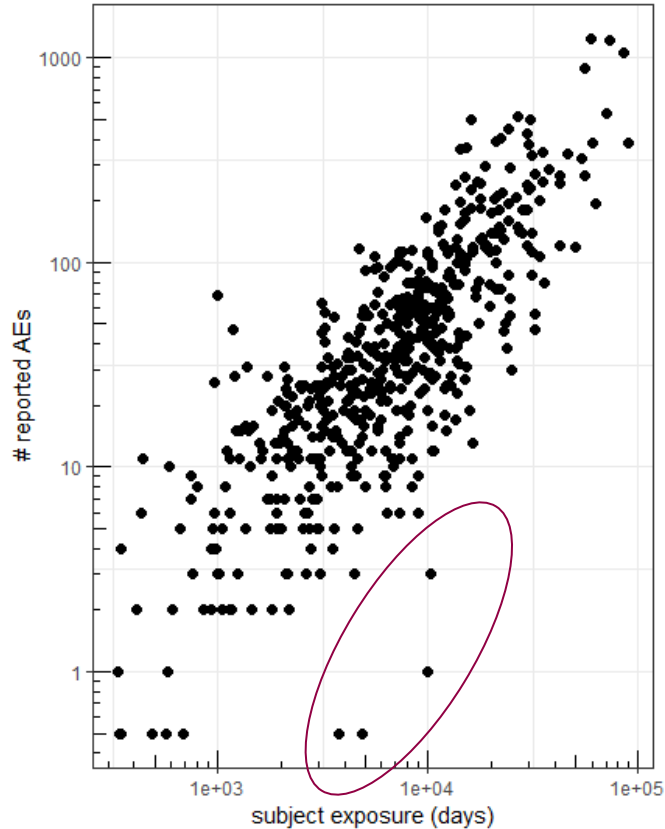


Statistical Counting Models: Poisson

- ▶ Poisson – models an expected number of counts in a fixed timespan
 - Observations over different timespans accounted for using an explicit “offset” covariate with parameter set to 1.0
 - Using R’s “glm” formula:

```
glm(AEcounts ~ 1 + offset(log(exposure)) ,  
    family="poisson")
```
- ▶ Poisson model assumes that variance is equal to mean rate
 - Can be relaxed using `family="quasipoisson"` but

the Poisson treatment: ~all sites significant!

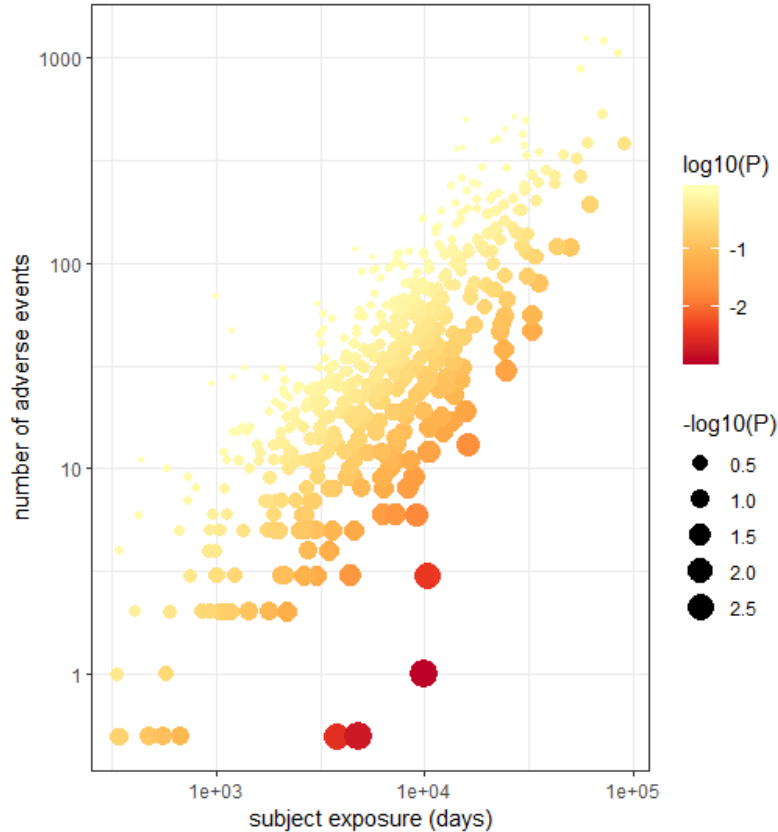
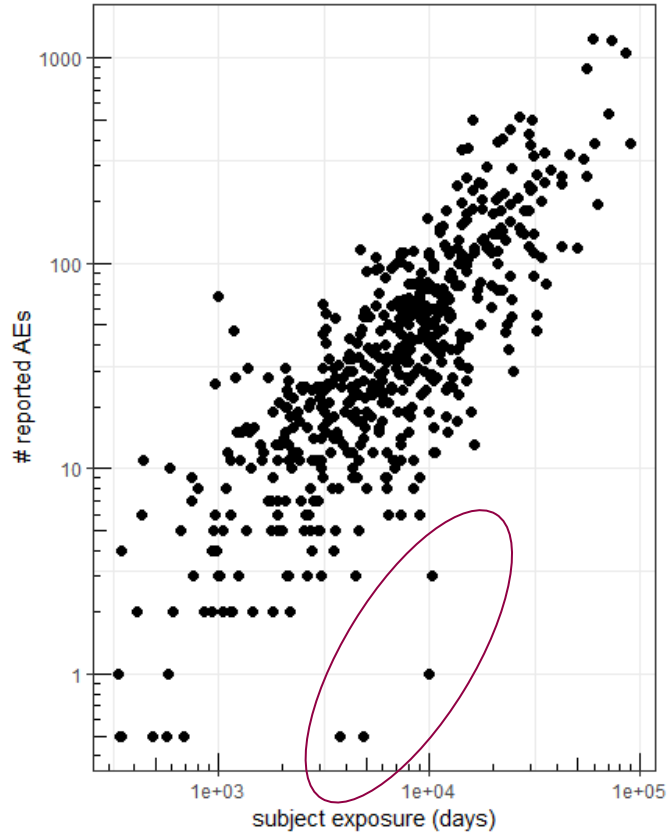


Statistical Counting Models: Negative Binomial

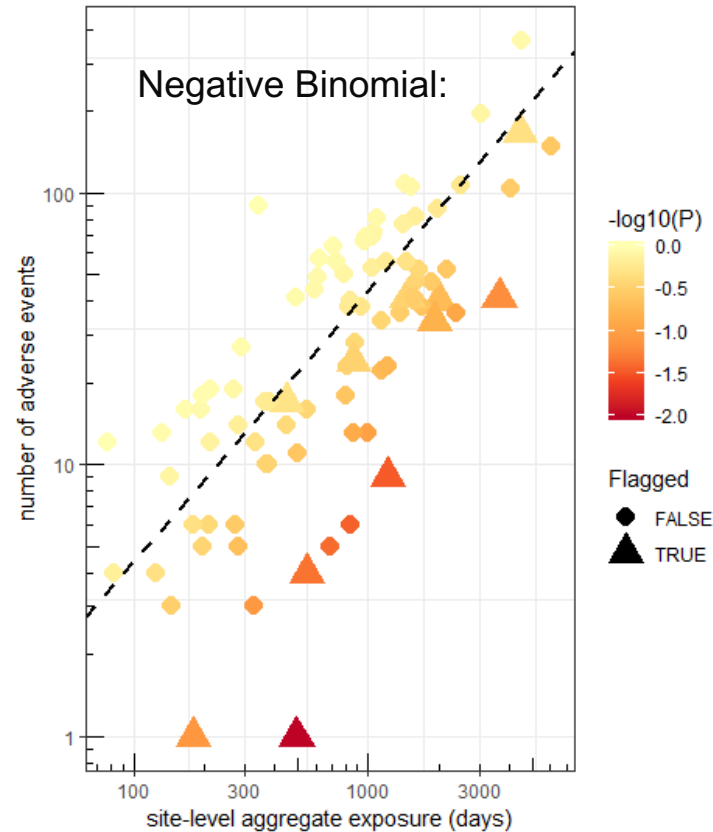
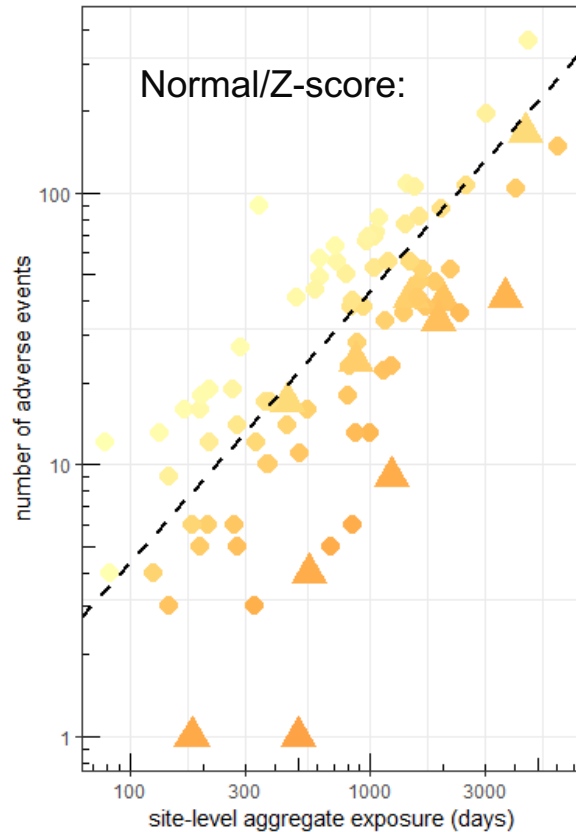
- ▶ Negative Binomial (NB) – as Poisson, models an expected number of counts in a fixed timespan, but with independent variance
 - Observations over different timespans accounted for using an explicit “offset” covariate with parameter set to 1.0
 - Using R’s “MASS” library and the “glm.nb” formula:

```
glm.nb(AEcounts ~ 1 + offset(log(exposure)))
```
- ▶ NB still assumes homogenous rates over time

the NB treatment: a few sites significant



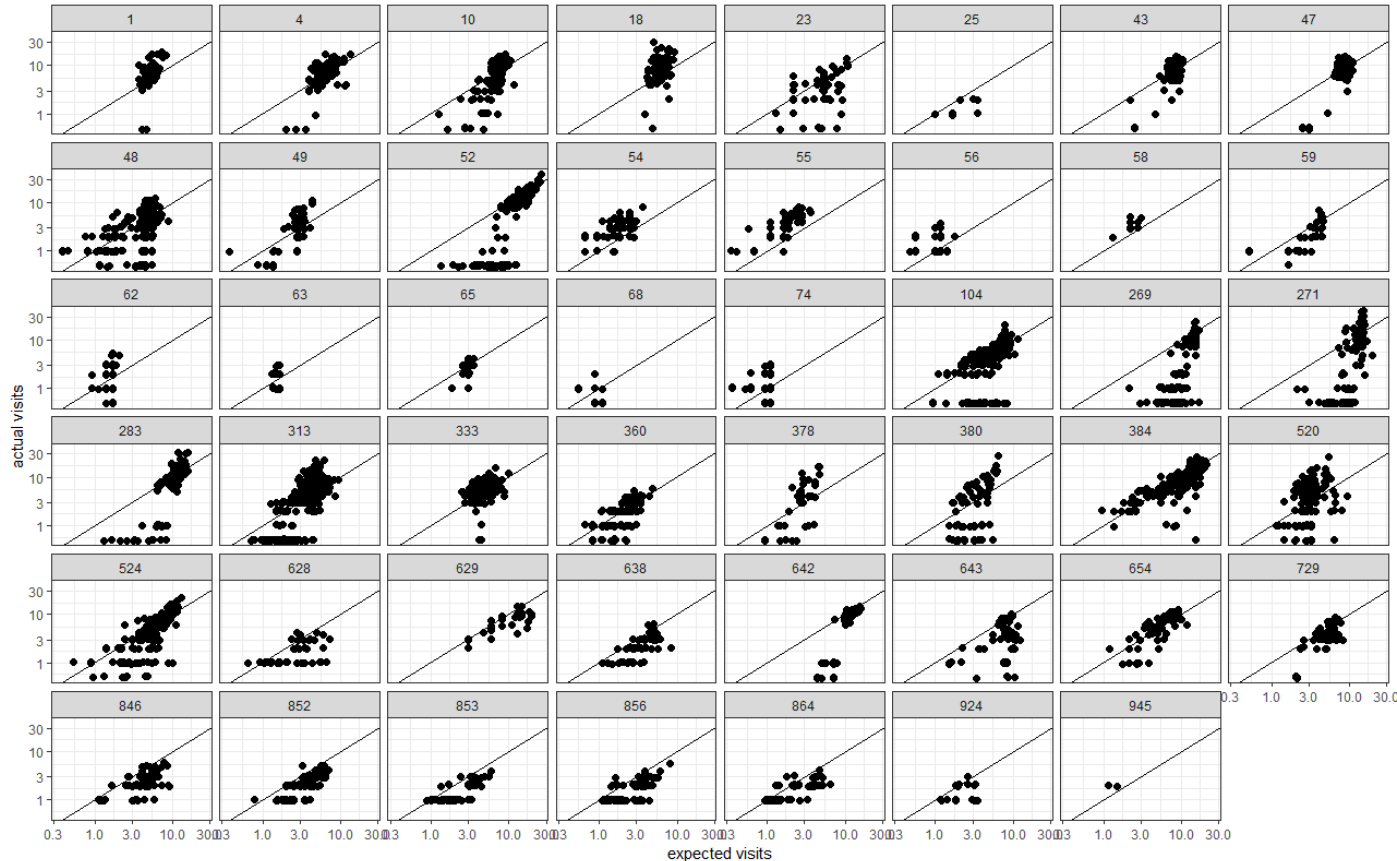
Sponsor-flagged AE under-reporting:



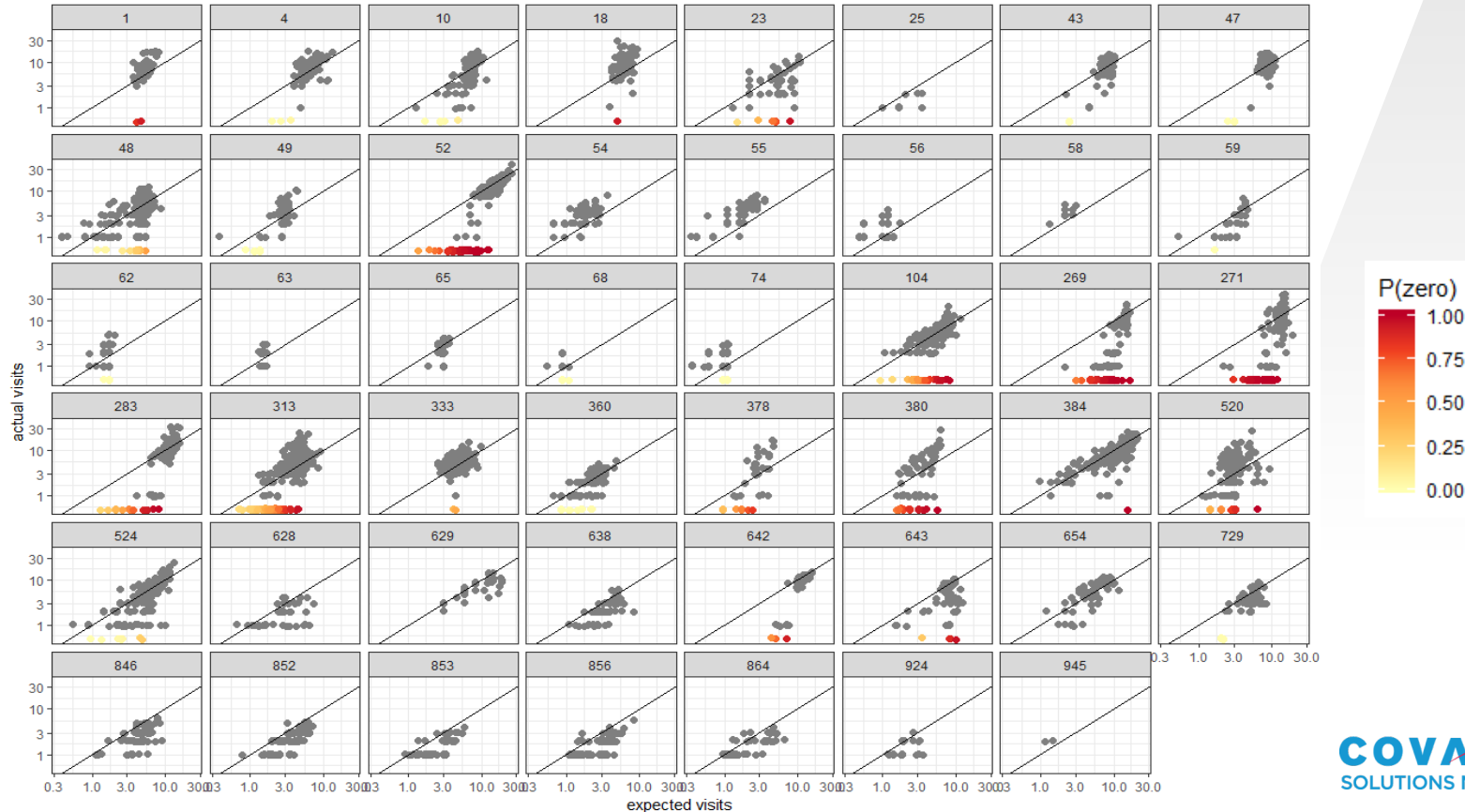
Statistical Counting Models: Zero-inflated

- ▶ observed zeroes could be structural, not by chance
 - Zero subjects enrolled in an activated site, yet site not actively enrolling in trial
 - Zero CRA monitoring visits at an active site, because no subjects enrolled
- ▶ both Poisson and Negative Binomial models can be extended as a mixture model with a non-zero fraction of structural zeroes mixed within P/NB-distributed random observations
 - Zero-inflated Poisson, or ZIP
 - Zero-inflated NB, or ZINB

CRA visit compliance: visit counts with zeroes



CRA visit compliance: visit counts with zeroes



In Summary

- ▶ Use counts, not rates, to ask statistical questions
 - better power, treatment of uncertainty
- ▶ Poisson commonly used, but beware variance == mean assumption!
- ▶ Negative Binomial convenient, flexible solution
- ▶ Zero-inflation (and other “hurdle” models) allow for mixtures of observations from different sources
- ▶ Further resources:

<https://cran.r-project.org/web/packages/pscl/vignettes/countreg.pdf>

<https://cran.r-project.org/web/packages/glmmTMB/vignettes/glmmTMB.pdf>



Covance Inc., headquartered in Princeton, NJ, USA, is the drug development business of Laboratory Corporation of America Holdings (LabCorp). COVANCE is a registered trademark and the marketing name for Covance Inc. and its subsidiaries around the world.