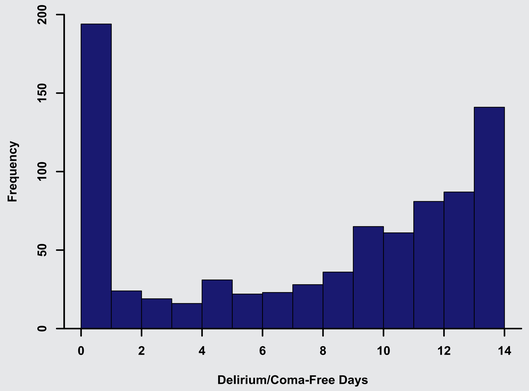
Power in the Case of Extreme Bimodal Data: Analysis of Delirium/Coma-Free Days in the ICU

Jennifer L. Thompson, MPH
and Ayumi Shintani, PhD
Vanderbilt University School of Medicine

Motivation

- Analysis of data in critical care medicine can be quite complex, partly due to unusual distribution of outcomes
- Common outcomes with this problem include event-free days, which are used to avoid bias due to death (for example, a patient who dies quickly has less opportunity to be on the ventilator than someone who survives the ICU, so using only length of time on the vent is problematic [1])
- Delirium/coma-free days: days a patient is alive and free of acute brain dysfunction over study period; highly clustered near zero and study endpoint [2]



Key Question:
Is there a regression method for these variables which is statistically appropriate, yet straightforward to interpret?

We examine the following methods:

- **Wilcoxon rank sum/Spearman correlation**
  no assumptions other than monotonic relationship, but unable to adjust for covariates
- **Traditional linear regression**
  straightforward interpretation, but data may not fit assumptions
- **Proportional odds logistic regression**
  not as sensitive to assumptions, but interpretation can be difficult [3]
- **Quantile regression**
  robust to outliers, but estimates of middle percentiles may be unstable
- **Negative binomial regression**
  able to handle cluster at 0, but may not be appropriate fit for second cluster [4]

Methods

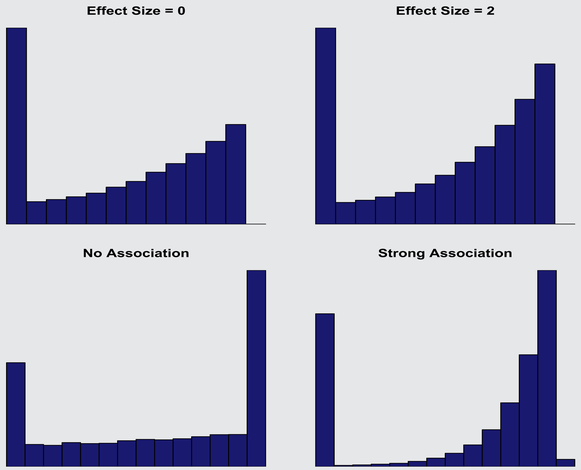
We simulated data with varying sample sizes and degrees of association between our outcome and either a dichotomous or continuous covariate (24 scenarios total).

Outcome

- Delirium/coma-free days (DCFDs) during a 14-day study period
- Used a combination of gamma, exponential and negative binomial distributions to simulate outcome under each scenario

Covariates

- **Dichotomous**: two groups with equal sample size; specified a given effect size (0, 1 or 2 DCFDs)
- **Continuous**: based on APACHE II scores at study enrollment; specified either no, moderate or strong association with DCFDs



Results

Dichotomous Covariate

Type I Error (Effect Size = 0)	50	100	200	600
Wilcoxon rank sum	2.0	4.7	5.8	12.0
Linear regression	2.3	4.9	5.6	12.3
Proportional odds	2.4	4.8	5.9	12.1
Quantile regression	3.2	2.9	4.5	8.1
Negative binomial	5.5	4.7	3.8	3.1

Effect Size = 1, 2 Days	50	100	200	600
Wilcoxon rank sum	6, 29	16, 63	32, 92	79, 100
Linear regression	6, 26	16, 63	33, 93	80, 100
Proportional odds	8, 30	17, 64	32, 92	79, 100
Quantile regression	7, 19	9, 38	20, 71	45, 97
Negative binomial	7, 11	8, 29	11, 71	47, 100

Continuous Covariate

Type I Error (No Association)	50	100	200	600
Spearman correlation	5.2	5.3	10.4	15.7
Linear regression	4.4	5.1	6.2	4.6
Proportional odds	5.8	5.8	10.3	14.8
Quantile regression	6.4	6.5	8.2	9.5
Negative binomial	5.1	5.4	6.3	6.5

Moderate, Strong Assoc.	50	100	200	600
Spearman correlation	7, 7	7, 10	9, 20	22, 43
Linear regression	18, 29	26, 76	48, 97	92, 100
Proportional odds	7, 6	6, 9	8, 18	22, 40
Quantile regression	5, 4	5, 9	6, 14	13, 16
Negative binomial	5, 6	7, 13	8, 19	13, 40

Conclusions

- Wilcoxon/Spearman, linear regression and proportional odds logistic regression tend to have very similar power to detect the same effect size with dichotomous covariates
- Quantile and negative binomial regression are not as powerful for outcomes with this distribution, especially with smaller sample size
- **Linear regression** seems to be the best choice for detecting associations between outcome and either type of covariate

References

1. Schoenfeld DA, Bernard GR; ARDS Network. Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials oftreatments for acute respiratory distress syndrome. Crit Care Med. 2002 Aug;30(8)1772-7.
2. Pandharipande PP et al. Effect of Sedation with Dexmedetomidine vs. Lorazepam on Acute Brain Dysfunction in Mechanically Ventilated Patients: The MENDS Randomized Controlled Trial. JAMA. 2007;298(22):2644-2653.
3. McCullagh, P. (1980). Regression models for ordinal data (with discussion). Journal of the Royal Statistical Society Ser. B, 42, 109-142.
4. Cameron AC and PK Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, 1998.

SCHOOL OF MEDICINE
VANDERBILT UNIVERSITY