A Bayesian Multivariate Longitudinal Model for Number of Sex Partners and HIV Status Disclosure

Robert Weiss and Lily Altstein

robweiss@ucla.edu

UCLA Biostatistics

Overview

- Project CLEAR
- Multivariate Longitudinal Data
- Multivariate Modeling with Random Effects
- Bayesian Methods
- Results
- Critique

Project CLEAR

- HIV+ drug-using youth 16-29
- Choose Life, Empowerment, Action, Results
  - Intervention: 18 90 minute weekly sessions;
  - (Phone or in-person) or delayed/control
  - Physical health/ARV/Disclosure
  - Unprotected sex acts
  - Emotional Health/ QOL
- LA, NY & SF
- 78% Male
- 55% High School Grad
- 75% Gay/Bisexual/Questioning

Data Set Description

- 173 subjects with usually 4 to 5 observations
- 70% MSM
- 69% in interventions
- On average 1 named partner, 3 unnamed partners
- 55% no unnamed partners
- 63% disclosure to named partners
- 55% disclosure to unnamed partners
- Partner is HIV+? 46% named, 35% unnamed

Goals

- General statistical methodology to model
  - Self-reported sexual and drug behavior
  - Longitudinally collected
  - Nesting: Disclosure within partners within time point
  - Multivariate outcomes: # partners, drug use, disclosure, acts
- Illustrate potential
- Many data sets to apply methods to
  - TLC, CLEAR, HLP
  - SATHCAP, AEIDRP, MultiMates

Complex Nested Longitudinal Data

- Hierarchical Data
  - Subjects assessed repeatedly over time
  - Sex partners within a subject at a given time
  - Sex acts within partner
  - Condom use within an act
  - HIV status disclosure to partner
  - HIV status of partner
- Drug use
  - By subject
  - Across time
  - With a partner, with specific sex act
- Data may be aggregated or broken out by act

Data: Qualitative

- Observations on subject i over time j
- Subjects $i = 1$ up to n
- j =time: baseline is $j = 0$. Then $j = 3, 6, 9$, and 15 months
- At each time: # sex partners
  - Named partners l
  - Unnamed/anonymous partners
- Named: Disclosure of HIV status by subject to each partner
- Unnamed: Global fraction disclosed to

Data Quantitative

- Count Y_{ijk},
  - $i =$ subject, $j =$time
  - $k = 1$ # named partners
  - $k = 2$ # unnamed/anonymous partners
- W_{ij1l} Disclosure?
  - Yes/no to named partner l
  - W_{ij2} fraction disclosed to unnamed partners

Modeling: 5 Outcomes

- 1. Counts of named partners: Poisson
- Counts of unnamed partners: Zero Inflated Poisson
  - 2. Has subject zero unnamed partners? Yes/No - Bernoulli
  - 3. If subject has at this time, how many? Poisson
- Disclosure to
  - 4. Specific named partner: yes/no - Bernoulli
  - 5. Globally to unnamed partners: fraction - Binomial

Population Effects: Time, Intervention

For all 5 outcomes:

- Time trend
  - Intercept and slope
  - Change in slope at month 3
- Equivalent to
  - Separate mean at baseline
  - Plus independent linear trend starting with month 3
- Separately for immediate treatment and delayed treatment groups
- Treatment: combined phone and in-person groups

Additional Population Covariates

- Count of named: Subject is MSM yes/no
- Has unnamed: MSM
- Count of unnamed: MSM
- Disclosure to named: MSM, Is unnamed> 0?, Count of Named
- Disclosure to unnamed: MSM, Count of Named, % HIV+ partners

Random Effects

- Subject specific random intercept for each of
  - Count of named
  - Has unnamed?
  - Count of unnamed
  - Disclosure to named
  - Disclosure to unnamed
- All correlated: 5-variate random effect
- Disclosure random intercept nested in subject*time
  - Subject * time disclosure to named

Purpose of Random Effects (REs)

- Observations within subject are correlated
- RE induces correlation in the modeling.
- Example: Outcome k (subject i, time j) = predictors(i, j, k)*coefficients(k) + random effect(ik)
- Same RE for multiple observations $j = 1, \dots J$
- Over-dispersion: Poisson and Binomial variance is function of a mean. Random effects de-link the mean and the variance, increasing the variance for a given mean.
- Nested subject*time REs for disclosure to named
  - Increases correlation within subject i's named partners at time j.
  - Decrease correlation within subject across times

Bayesian Methods: Disadvantages

- Unfamiliar: Requires education of
  - Statisticians
  - Colleagues
  - Editors
  - Referees
- Requires public relations
- Different computational issues
  - Can take longer to run
  - Not all issues negative
  - Can fit more data sets & models
- All assumptions are explicit

Bayesian Methods: Advantages

- Can include outside / additional information (prior information)
- Can fit sparse data sets, small sample sizes
- More stable estimation: fewer wild results
- Complex models can be built
  - Greater fidelity to underlying scientific theory
  - Complex hypotheses to be studied
  - Greater fidelity to data collection
- Scientifically more satisfying statistical conclusions
  - Probability alternative hypothesis is true.
  - Probability treatment effect is positive.
- More complex conclusions possible.

Priors: The Effect of MSM

- Prior for a coefficient is an assumption.
- Counts of named and unnamed partners.
- Counts are positive. Coefficients on log count scale.
- Prior for MSM normal, mean zero, sd 2, N$(0, 2^2)$.
- Mean zero: A priori balanced pos vs. neg.
- $\exp(-1.96 * 2) \approx .02$, $\exp(1.96 * 2) \approx 50$.
- Prior 95% interval: MSM has .02 to 50 times as many partners as non-MSM.
- Better: MSM have .1 to 10 times as many N$(0, 1.15^2)$.
- $\log(10) = 2.30$, assume $\log(.1)$ to $\log(10)$ is 95% interval.
- Even better? .5 to 2 times as many partners, N$(0, .35^2)$.

Effect of MSM 1

Count of Named Partners

Prior normal, mean zero, sd 2.

- A posteriori result: a .95% interval is MSM have 1.06 to 1.83 times as many partners as non-MSM.
- Probability that the MSM effect is positive is .99.
- One-sided p-value .01.
- Can imbed model in larger model
  - A priori prob of null (MSM = non-MSM) is .5.
  - Alternative hypothesis is coefficient N$(0, 2^2)$.
  - Posterior probability of null hypothesis is .48.
  - Better prior assumptions give $(.3, .2)$.

Priors and Effect of MSM 2

Non-named Partners *IF* have non-named partners

- Result: MSM have from 5 to 25 times as many unnamed partners as non-MSM, with probability .95.
- One sided p-value is .000001.
- Imbedded model: Probability of null hypothesis: .00005.
- Overwhelming information.

Priors and Effect of MSM 3

Probability of Having Unnamed Partners.

- Prior is .95 probability of odds ratio being in $(.14, 7)$
- One sided P-value is .2.
- Probability that effect is positive is .8.
- Probability that null hypothesis is true is .52.
- Posterior interval is .4 to 8 MSM odds of having unnamed partners over non-MSM.
- No information in data on this effect.

Random Effects Correlations

Correlations among average level of variables

- Positive and significant correlation .37 to .46 among
  - Counts of named partners
  - Probably of having unnamed partners
  - Counts of unnamed partners
- Positive and significant correlation .68 among
  - Disclosure to named partners
  - Disclosure to unnamed partners
- Negative non-significant cross correlations

Additional Effects: Summary

- No identifiable treatment effect, except possibly on disclosure to unnamed.
- MSM: no effect on disclosure
- Partner being HIV+ has major positive effect on disclosure for both named and unnamed.
- Having unnamed partners doesn't affect disclosure to named partners.
- Count of named (unnamed) has no apparent effect on disclosure to unnamed (named).

Breaking News: New Model Run

Treatment effects removed. Improved prior for MSM for counts. No redundant predictors (# partners predicting disclosure)

- **# named partners RE significantly correlated with less disclosure to named partners RE.** $\rho = -.3$.
- # Named partners sig. corr. # unnamed partners $\rho = .4$
- Named disclosure sig. corr. unnamed disclosure $\rho = .4$
- MSM **NOT sig** predictor for # named partners $(.95, 1.6)$
- MSM **sig** predictor of having unnamed partners $(2.3, 16)$
- MSM **sig** predictor of # unnamed partners $(1.01, 3.7)$
- Partner HIV+ is major predictor of disclosure $(1.4, 3.7)$

Statistical Conclusions

- Need convenient and communicable conclusions for non-statisticians.
- Several choices for Bayesian conclusions.
  - Mean, sd
  - Confidence interval (CI) on logit or log scale
  - Transformed CI - odds or mean multiplier
  - Probability of null (alternative) hypothesis is true
  - Probability effect is positive (negative).
- Tune priors (incorporate expert opinion).
- Cross level predictors: raw data, fitted values, REs.
- For subject matter paper, choose covariates of interest.