# Addressing construct invalidity in partial interval recording data

James E. Pustejovsky
UT Austin
pusto@austin.utexas.edu

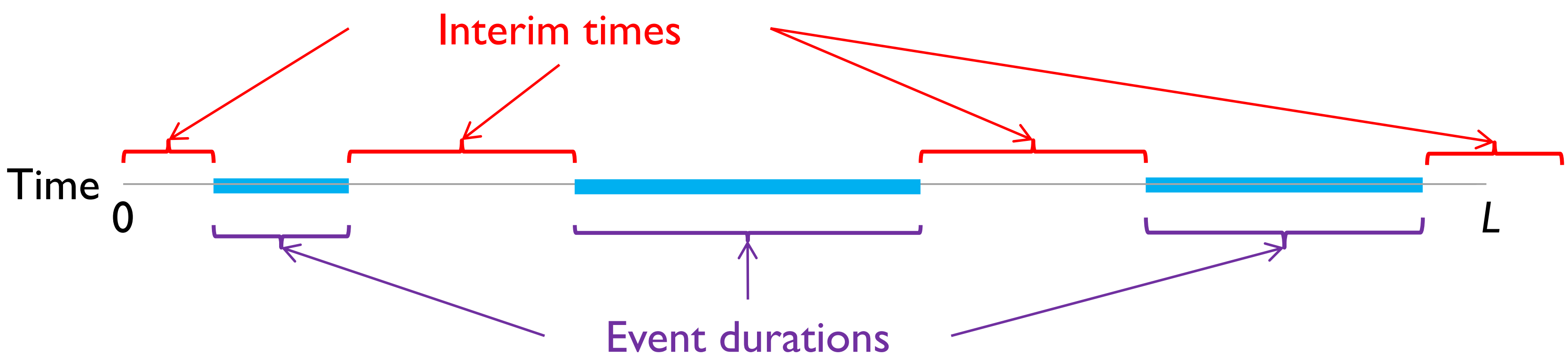
Thanks to collaborators:
Daniel Swan
Christopher Runyon

Texas Universities Education Statistics and Psychometrics Meeting
March 21, 2014

# Direct observation of behavior

- Applications in psychology and education research
  - Assessing teaching practice
  - Measuring student behavior
  - Evaluating interventions for individuals with disabilities

# The observed behavior stream

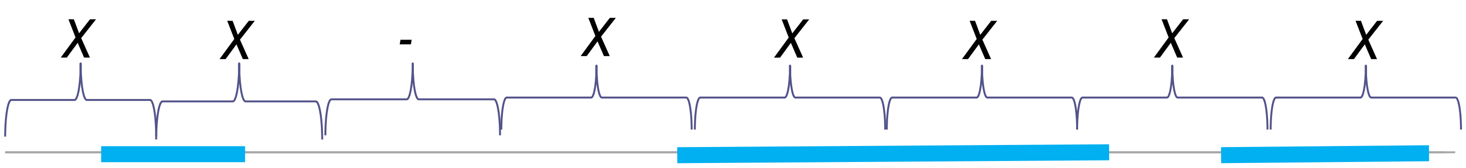



- Behavioral characteristics of interest
  - **Prevalence**: proportion of time that a behavior occurs
  - **Incidence**: rate at which new behaviors begin

# Partial interval recording (PIR)

- Divide observation time into $K$ intervals, each of length $c$.
- For each interval, record whether behavior occurred ***at any point*** during the interval.
- Calculate the proportion of intervals with behavior:

$Y^{PIR}$ = (# Intervals with behavior) / $K$.



# A model for observed behavior

## Equilibrium Alternating Renewal Process
**(Rogosa & Ghandour, 1991)**

1. Event durations are identically distributed random variables, with mean duration $\mu > 0$.
2. Interim times (ITs) are identically distributed random variables, with mean IT $\lambda > 0$.
3. Event durations and ITs are all mutually independent.
4. Process is aperiodic and in equilibrium.

## Target parameters

- Prevalence (phi): $\phi = \frac{\mu}{\mu + \lambda}$ Incidence (zeta): $\zeta = \frac{1}{\mu + \lambda}$

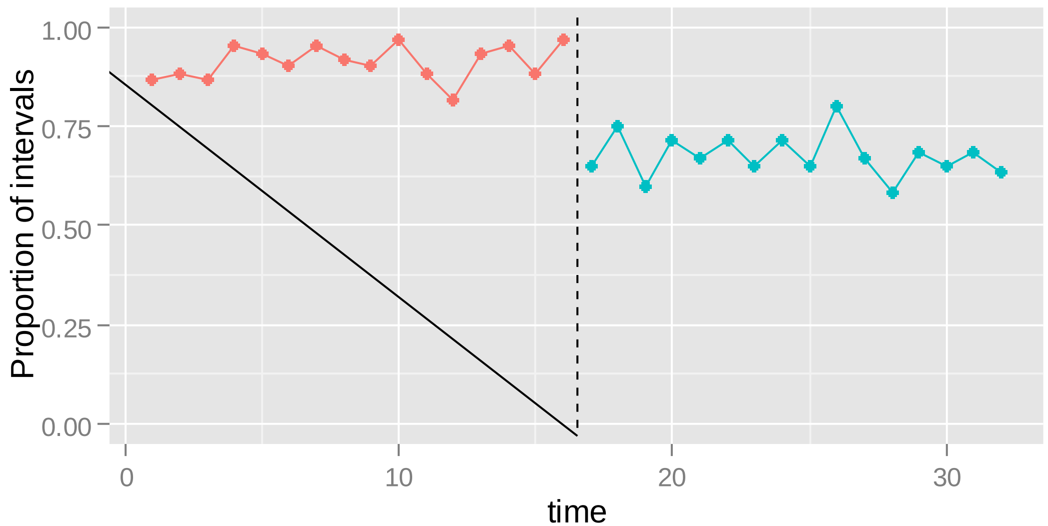
# Construct invalidity of PIR

- Under an equilibrium alternating renewal process,

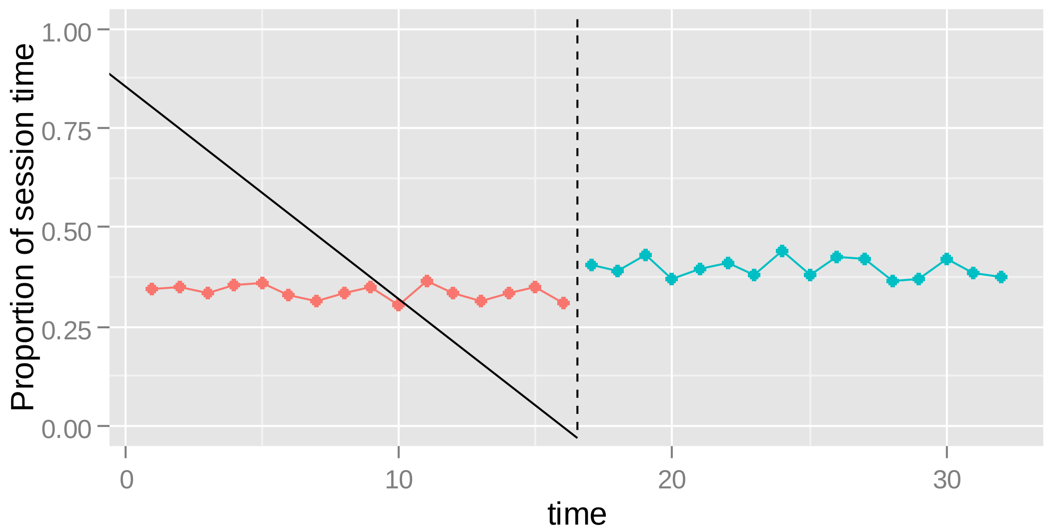
$$E\left(Y^{PIR}\right) = \phi + \zeta \int_0^c \Pr\left(IT > x\right) dx$$

# Potential for misleading inferences

Using PIR, it appears that prevalence decreases…




…when sample prevalence has instead increased slightly.

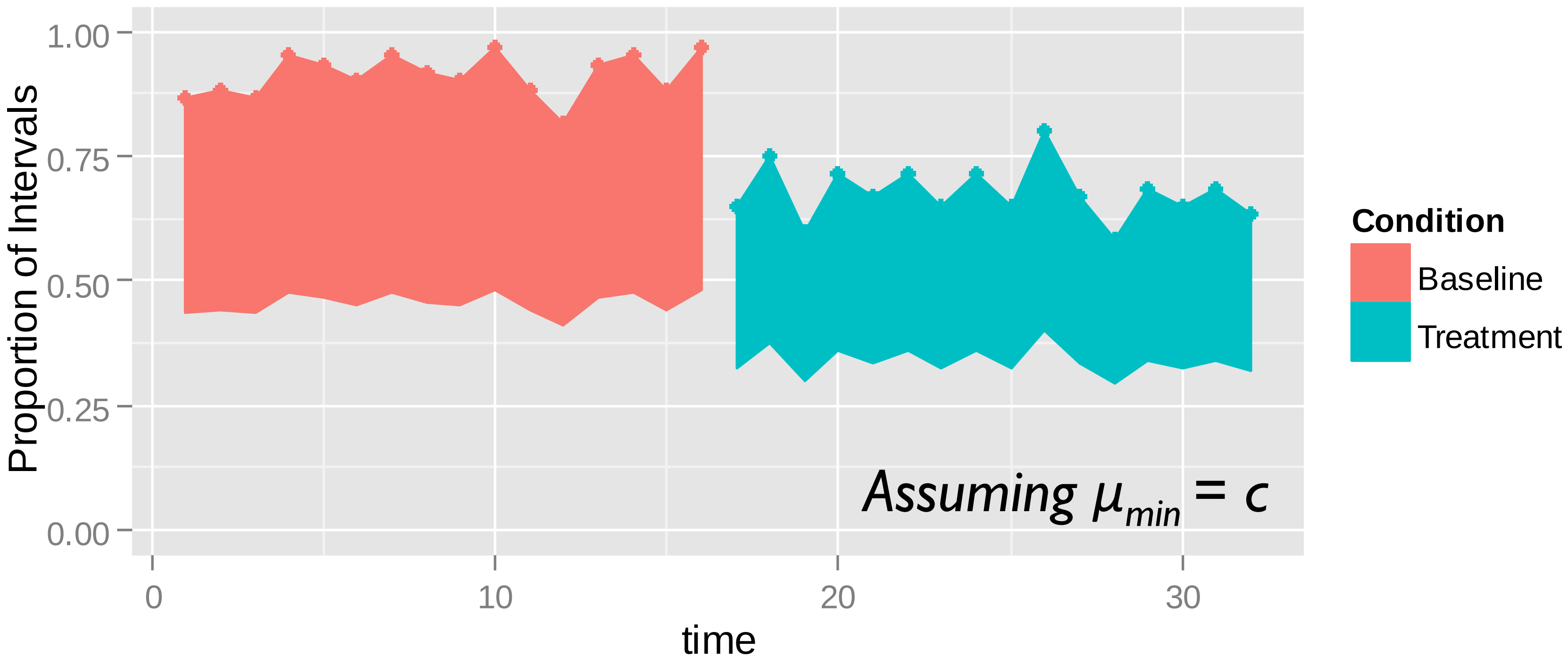


# Methods to address construct invalidity

1. If mean event duration is long
   - bound for prevalence
2. If mean event duration is short, interim times are mostly long
   - bound for incidence
3. If event durations and interim times are exponential
   - moment estimators for prevalence and incidence
4. If you can get interval-level data
   - Likelihood estimators for prevalence and incidence

# Method 1: A bound for prevalence

- Pick a value $\mu_{min}$ where you are certain that $\mu > \mu_{min}$
- It follows that $\frac{\mu_{min}}{\mu_{min} + c} E\left(Y^{PIR}\right) < \phi < E\left(Y^{PIR}\right)$



# Discussion

- Methods all involve fairly strong assumptions
- Prospectively, choose interval length based on what you know about the behavior.
- Use partial interval recording only if one of the analytic methods can be justified a priori.

# Method 2: A bound for incidence

- Pick a value $\mu_{max}$ where you are certain that $\mu < \mu_{max}$
- Pick a value $p$ where you are certain that Pr(IT > $c$) > $p$.

- It follows that

$$\frac{E\left(Y^{PIR}\right)}{\mu_{\max}+c} < \phi < \frac{E\left(Y^{PIR}\right)}{(1-p)c}$$

# Method 3: A bound for interim time ratio

- Two samples of PIR data, $s = 0, 1$, generated from ARP($\mu_s$, $\lambda_s$).
- Assume that $\mu_0 = \mu_1$.
- Assume that interim times are exponentially distributed.
- If $E(Y_0) < E(Y_1)$, then

$$\text{cll}\left[E\left(Y_1\right)\right] - \text{cll}\left[E\left(Y_0\right)\right] < \ln\left(\frac{\lambda_0}{\lambda_1}\right) < \text{logit}\left[E\left(Y_1\right)\right] - \text{logit}\left[E\left(Y_0\right)\right]$$

$$\text{cll}(x) = \ln\left(-\ln(1-x)\right)$$

$$\text{logit(x)} = \ln(\text{x}) - \ln(1-\text{x})$$

# Method 4: Moment estimators

- Assume that event durations and interim times are both exponentially distributed (i.e., Alternating Poisson Process).

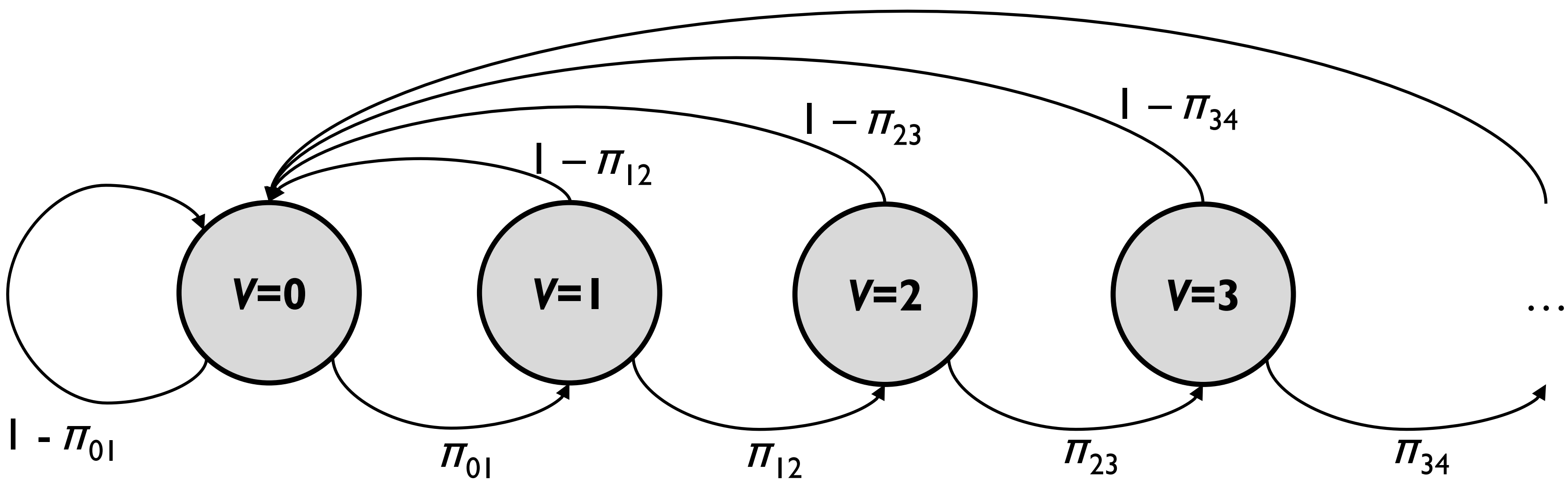
$$E\left(Y\right)=1-(1-\phi)\exp\left(\frac{-\zeta c}{(1-\phi)}\right)$$

$$Var\left(Y\right)=\frac{E(Y)\left[1-E(Y)\right]}{K}\left[1+\frac{2\phi}{KE(Y)}\sum_{k=1}^{K-1}(K-k)\exp\left(\frac{\zeta c}{\phi}-\frac{\zeta kL}{\phi(1-\phi)K}\right)\right]$$

- Replace expectation and variance with sample mean and sample variance, solve for $\phi$ and $\zeta$.
- Bootstrap confidence intervals.

# Method 5: Likelihood estimators with interval-level data

- Define $V_k$ as the number of consecutive intervals where behavior is present, $k = 1,\ldots,K$.
- Under the Alternating Poisson Process, $V_1,\ldots,V_K$ follow a Discrete Time Markov Chain on the space {0,1,2,3,…}.



# DTMC model for PIR intervals

Transition probabilities:

$$\pi_{j,j+1} = \Pr\left(V_k = j+1 \mid V_{k-1} = j\right) = 1 - e^{-L/\lambda}\left[1 - f^{(j)}(0)\right]$$

where

$$f(q) = \frac{\phi - (\phi - q)\exp\left[\dfrac{-L\zeta}{\phi(1-\phi)}\right]}{1-(1-q)\exp\left[\dfrac{-L\zeta}{(1-\phi)}\right]}$$

and $f^{(j)}$ is the $j$-fold recursion of $f$.